análise léxica

converte sequência de caracteres em sequência de elementos lexicais (tokens)

< tokenname, value >

3.1 Linguagens, gramáticas e expressões regulares

Linguagem regular

$\emptyset$ é uma LR

$\forall a \in A$ (alfabeto), o conjunto $\{a\}$ é uma LR elementos primitivos

L1, L2 linguagens regulares, então

- $L1 \cup L2$ é LR união $\{ab\} \cup \{bc, c\} = \{ab, bc, c\}$
- $L1 . L2$ é LR concatenação $\{ab\} . \{bc, c\} = \{abbc, abc\}$
- $(L1)^*$ é LR fecho $\{ab\}^* = \{\varepsilon, ab, abab, ababab, \dots\}$

fecho de kleene

$L^* = L_1 \cup L_2 \cup \dots$

$= \bigcup_{i=0}^{\infty} L_i$

$= \{\varepsilon\} \cup L$

expressões regulares por indução

() é ER que representa a LR {}

$\forall a \in A$, a é ER que representa a LR $\{a\}$

se e1 e e2 são ER que representam as LR L1 e L2

- então $(e_1 | e_2)$ é ER representando a LR $L1 \cup L2$ escolha
- então $(e_1 e_2)$ é ER representando a LR $L1 . L2$
- então e_1^* é ER representando $L1^*$

$ER()^* =_{LR} \{\varepsilon\}$

Propriedades ER

escolha |

- comutativa $e_1 | e_2 = e_2 | e_1$
- associativa $e_1 | (e_2 | e_3) = (e_1 | e_2) | e_3 = e_1 | e_2 | e_3$
- idempotência $e_1 | e_1 = e_1$
- elemento neutro $e_1 | () = e_1$

concatenação •

- associativa $e_1 . (e_2 . e_3) = e_1 . e_2 . e_3$
- elemento neutro $e_1 . \varepsilon = e_1$
- elemento neutro $e_1 . () = ()$

mistas | .

- distributiva (à esq e dir) $e_1 . (e_2 | e_3) = e_1 . e_2 \ | \ e_1 . e_3$

fecho *

$(e^*)^* = e^*$

$(e_1^* | e_2^*)^* = (e_1 | e_2)^*$

$(e_1 | e_2^*)^* = $

$= (e_1^* | e_2)^* = (e_1 | e_2)^*$

precedência

fecho concatenação escolha

EXERCÍCIOS

$L = \{w \in A^* : \#(a, w) = 2\}$

o nr de ocorrências do símbolo a em w é dois outras notações ER

$e^+ = e.e^*$ — $[a_1 a_2 \ldots a_n] = (a_1 | a_2 | \ldots | a_n)$

$e^? = (e|\varepsilon)$ — $[a_1 - a_n] = (a_1 | \ldots | a_n)$

$e\{n\}$ n ocorrências — $[\hat{}\, a_1 \ldots a_n]$ símbolo fora do conjunto dado

$e\{n_1, n_2\}$ entre n_1 e n_2 — $[\hat{}\, a_1 - a_n]$ " " " " "

$e\{n_1,\}$ n_1 ou + ocorrências

. símbolo qualquer

^ \$ palavra vazia no início/fim de linha

\< \> palavra vazia no início/fim de palavra

gramáticas regulares

os símbolos não terminais só podem aparecer no fim!

$S \to a\,X$ $\quad X \to a\,X \mid b\,X \mid \varepsilon$

$G = (T, N, P, S)$ tu não podes subir

T conjunto finito ñ vazio de símbolos terminais

N disjunto de T, ñ vazio de símbolos não terminais

P conjunto de produções (regras) na forma $\alpha \to \beta$

$\alpha \in N$

$\beta \in T^* \mid T^* N$

$S \in N$ é o símbolo inicial

Maiúsculas símbolos não terminais
minúsculas símbolos terminais
ao contrário do ANTLR

∴ linguagem gerada por uma gramática regular é regular

⇒ é possível converter gramática em ER que represente a LR e vice-versa

operações GR

GR são fechadas sobre operações de reunião, concatenação, fecho, interseção e complementação

reunião começar por obter GR para linguagens individuais!

$T = T_1 \cup T_2$

$N = N_1 \cup N_2 \cup \{S\}$

$P = \{S \to S_1, S \to S_2\} \cup P_1 \cup P_2$

S

concatenação produções T^* de S_1 ganham S_2 no fim!

$T = T_1 \cup T_2$

$N = N_1 \cup N_2$

P alterar S_1 para invocar S_2 ou introduzir $S \to S_1 S_2$

$S = S_1$

fecho produções T^* ganham $S \in N$ no fim

$T = T_1$

$N = N_1 \cup \{S\}$

$P = \{S \to \varepsilon, S \to S_1\}$ mas S_1 aponta para S ou introduzir $S \to \varepsilon | S_1 S$

S

---

conversão ER → GR

dadas as ER básicas e, e^*, $e_1 | e_2$, $e_1 e_2$
identificam-se as GR relativas às ER e fazem-se as operações sobre as gramáticas

> 1.º desconstruir ER em árvore
>
> 2.º construir GR a partir das folhas

e ε $S \to \varepsilon$

    a $S \to a$

e^* fecho da gramática equivalente a e

$e_1 | e_2$ reunião das GR

$e_1 e_2$ concatenação das GR

conversão GR → ER

converte-se $G = (T, N, P, S)$ em

$\mathcal{E} = \{(E, \varepsilon, S)\}$

$\cup \{(A, w, B) : (A \to wB) \in P \wedge B \in N\}$

$\cup \{(A, w, \varepsilon) : (A \to w) \in P \wedge w \in T^*\}$, com $E \notin N$

por transformações de equivalência, removem-se um a um os símbolos de N até obter (E, e, ε) sendo e a ER

Para remover símbolos de N

substituir (A, β_i, B) por (A, ω, B), $\omega = \beta_1 | \beta_2 | \ldots | \beta_n$

substituir $(A, w_1, B), (B, w_2, B), (B, w_3, C)$

por $(A, w_1 w_2^* w_3, C)$