Universitetet i Oslo
Institutt for Informatikk
Reliable Systems
Martin Steffen

INF 5110: Compiler construction

Spring 2023 **Handout 4** 20. 2. 2023

Handout 4: Parsing

Issued: 20. 2. 2023

For reference, to follow the slides, the handout includes some grammars we repeteadly used for illustration. These are various versions of the context-free grammar for expressions. The first version is the "obvious" one. Also some definitions are added.

Some definitions

Definition 1 (First set) Given a grammar G and a non-terminal A. The *first-set* of A, written $\mathit{First}_G(A)$ is defined as

$$\mathit{First}_G(A) = \{a \mid A \Rightarrow_G^* a\alpha, \quad a \in \Sigma_T\} + \{\epsilon \mid A \Rightarrow_G^* \epsilon\} \; . \tag{1}$$

Definition 2 (Nullable) Given a grammar G. A non-terminal $A \in \Sigma_N$ is *nullable*, if $A \Rightarrow^* \epsilon$.

Definition 3 (First set of a symbol) Given a grammar G and grammar symbol X. The *first-set* of X, written $\mathit{First}(X)$, is defined as follows:

1. If $X \in \Sigma_T + \{\boldsymbol{\epsilon}\}$, then $\mathit{First}(X)$ contains X.
2. If $X \in \Sigma_N$: For each production

$$X \to X_1 X_2 \dots X_n$$

   (a) $\mathit{First}(X)$ contains $\mathit{First}(X_1) \setminus \{\boldsymbol{\epsilon}\}$
   (b) If, for some $i < n$, *all* $\mathit{First}(X_1), \dots, \mathit{First}(X_i)$ contain $\boldsymbol{\epsilon}$, then $\mathit{First}(X)$ contains $\mathit{First}(X_{i+1}) \setminus \{\boldsymbol{\epsilon}\}$.
   (c) If all $\mathit{First}(X_1), \dots, \mathit{First}(X_n)$ contain $\boldsymbol{\epsilon}$, then $\mathit{First}(X)$ contains $\{\boldsymbol{\epsilon}\}$.

Definition 4 (First set of a word) Given a grammar G and word α. The *first-set* of

$$\alpha = X_1 \dots X_n \; ,$$

written $\mathit{First}(\alpha)$ is satisfies the following conditions

1. $\mathit{First}(\alpha)$ contains $\mathit{First}(X_1) \setminus \{\boldsymbol{\epsilon}\}$
2. for each $i = 2, \dots n$, if $\mathit{First}(X_k)$ contains $\boldsymbol{\epsilon}$ for *all* $k = 1, \dots, i-1$, then $\mathit{First}(\alpha)$ contains $\mathit{First}(X_i) \setminus \{\boldsymbol{\epsilon}\}$

3. If all $\mathit{First}(X_1), \ldots, \mathit{First}(X_n)$ contain $\boldsymbol{\epsilon}$, then $\mathit{First}(X)$ contains $\{\boldsymbol{\epsilon}\}$.

Definition 5 (Follow set) Given a grammar G with start symbol S, and a non-terminal A. The *follow-set* of A, written $\mathit{Follow}_G(A)$, is

$$\mathit{Follow}_G(A) = \{a \mid S\,\$ \Rightarrow_G^* \alpha_1 A a \alpha_2, \quad a \in \Sigma_T + \{\,\$\,\}\} \ . \tag{2}$$

Definition 6 (Follow set of a non-terminal) Given a grammar G and nonterminal A. The *Follow-set* of A, written $\mathit{Follow}(A)$ is defined as follows:

1. If A is the start symbol, then $\mathit{Follow}(A)$ contains **$**.
2. If there is a production $B \to \alpha A \beta$, then $\mathit{Follow}(A)$ contains $\mathit{First}(\beta) \setminus \{\boldsymbol{\epsilon}\}$.
3. If there is a production $B \to \alpha A \beta$ such that $\boldsymbol{\epsilon} \in \mathit{First}(\beta)$, then $\mathit{Follow}(A)$ contains $\mathit{Follow}(B)$.

Lemma 7 (LL(1) (without nullable symbols)) A reduced context-free grammar without nullable non-terminals is an LL(1)-grammar iff for all non-terminals A and for all pairs of productions $A \to \alpha_1$ and $A \to \alpha_2$ with $\alpha_1 \neq \alpha_2$:

$$\mathit{First}_1(\alpha_1) \cap \mathit{First}_1(\alpha_2) = \emptyset \ .$$

Lemma 8 (LL(1)) A reduced context-free grammar is an LL(1)-grammar iff for all non-terminals A and for all pairs of productions $A \to \alpha_1$ and $A \to \alpha_2$ with $\alpha_1 \neq \alpha_2$:

$$\mathit{First}_1(\alpha_1 \mathit{Follow}_1(A)) \cap \mathit{First}_1(\alpha_2 \mathit{Follow}_1(A)) = \emptyset \ .$$

Definition 9 (Handle) Assume $S \Rightarrow_r^* \alpha A w \Rightarrow_r \alpha \beta w$. A production $A \to \beta$ at position k following α is a *handle of* $\alpha\beta w$. We write $\langle A \to \beta, k\rangle$ for such a handle.

References

[Engelfriet and Vogler, 1987] Engelfriet, J. and Vogler, H. (1987). Look-ahead on pushdowns. *Information and Computation*, 73(3):245–279.

[Hoogeboom and Engelfriet, 2004] Hoogeboom, H. J. and Engelfriet, J. (2004). Pushdown automata. In Martín-Vide, C., Victor, M., , and Păun, G., editors, *Formal Languages and Applications*, pages 117–138. Springer Verlag, Berlin, Heidelberg.