

프로그래밍언어론

Chapter 3 어휘 및 구문 분석

Syntactic sugar causes cancer of the semicolon.

구문설탕은 세미콜론 암을 일으킨다

A. Perlis

목차 Contents

3.1 촘스키 분류 Chomsky Hierarchy

3.2 어휘 분석 Lexical Analysis

3.3 구문 분석 Syntactic Analysis

3.1 촘스키 분류

- 정규 문법 Regular grammar -- least powerful
- 문맥자유 문법 Context-free grammar (BNF)
- 문맥민감 문법 Context-sensitive grammar
- 무제한 문법 Unrestricted grammar

정규 문법 Regular Grammar

- 가장 간단하고 가장 표현력이 약하다
- 다음과 동일한 표현력을 가진다:
 - 정규식 Regular expression
 - 유한상태 오토마타 Finite-state automaton
- 오른쪽 정규 문법 Right regular grammar:
(여기서 $\omega \in T^*$, $B \in N$)

$$A \rightarrow \omega B$$

$$A \rightarrow \omega$$

예제

- $Integer \rightarrow 0\ Integer \mid 1\ Integer \mid \dots \mid 9\ Integer$
 $\mid 0 \mid 1 \mid \dots \mid 9$

정규 문법 Regular Grammars

- 왼쪽 정규 문법 : 오른쪽과 동일
- 토큰 생성기를 구현할 때 사용한다.
- 문맥자유 문법보다 표현력이 약하다.
- 다음은 정규언어가 아니다:

$$\{ a^n b^n \mid n \geq 1 \}$$

쌍을 이루는 구조를 표현할 수 없다: (), { }, begin end

문맥자유 문법 Context-free Grammars

- BNF 표기법 : CFG를 표기하는 방법
- 푸쉬다운 오토마타의 표현력과 동일
- 모호성이 없는 많은 CFG에 대하여
 - 테이블 구동 방식의 파서를
 - 선형 시간 복잡도로 구현 가능

문맥민감 문법 Context-Sensitive Grammars

- 생성규칙:
 - $\alpha \Rightarrow \beta \quad |\alpha| \leq |\beta|$
 - $\alpha, \beta \in (N \cup T)^*$
- 좌변이 터미널과 논터미널들의 기호열로 구성

문맥민감 문법의 결정불가능 성질

- 스트링 ω 와 문법 G 가 주어졌을 때:
 - $\omega \in L(G)$?
 - $L(G)$ is non-empty ?

- 결정불가능(*Undecidable*)

주어진 문제와 임의의 입력에 대하여 해답을 가지고 멈추는 프로그램을 작성할 수 없음

무제한 문법 Unrestricted Grammar

- 동일한 능력(표현력, 계산능력)을 가지는 것:
 - Turing machine
 - von Neumann machine
 - C++, Java
- 모든 계산가능 함수를 계산할 수 있다.