

컴파일러: 2장

국민대학교 컴퓨터공학부
강 승 식

1

제2장 언어와 문법

- 문법이란? 스트링 = 문장
 - 문장들의 집합을 기술
 - 형식 언어에서 '문장'은 '스트링(string)'이라고 함
- 예1) 언어 $L_1 = \{ 00, 01, 10, 11 \}$ 에 대한 문법

$$S \rightarrow 00 \mid 01 \mid 10 \mid 11$$

$$\begin{aligned} S &\rightarrow 00 \\ S &\rightarrow 01 \\ S &\rightarrow 10 \\ S &\rightarrow 11 \end{aligned}$$

2

- 예2) 언어 $L_2 = \{ a, b, aa, ab, ba, bb \}$ 에 대한 문법

$$S \rightarrow a \mid b \mid aa \mid ab \mid ba \mid bb \quad \text{정규언어}$$

- 예3) 예2와 동일한 언어를 기술하는 문법

$$\begin{aligned} S &\rightarrow a \mid b \mid aX \mid bX \\ X &\rightarrow a \mid b \end{aligned} \quad \text{정규문법 (우선형)}$$

- 예4) 예2, 예3과 동일한 언어를 기술하는 문법

$$\begin{aligned} S &\rightarrow XY \mid X \\ X &\rightarrow a \mid b \\ Y &\rightarrow a \mid b \end{aligned} \quad \begin{array}{l} \text{같은 문법이지만} \\ \text{정규문법 } x \\ \text{이유는 } XY \mid X \text{ 때문 nonterminal이} \\ \text{context free grammar임} \end{array}$$

- 예5) 언어 $L_3 = \{ 0, 1, 00, 01, 10, 11, 000, 001, \dots \}$

정규문법 x nonterminal 두개 이상(XS)

$$\begin{aligned} S &\rightarrow XS \mid 0 \mid 1 \\ X &\rightarrow 0 \mid 1 \end{aligned}$$

- 예6) 잘못된 문법의 예

$$\begin{aligned} S &\rightarrow XS \\ X &\rightarrow a \mid b \end{aligned}$$

이유는 XY | X 때문 nonterminal이 두개 있어서 (XY)

4

문법의 정의

- 문법의 4가지 구성 요소
 $G = (T, N, P, S)$

- T: 터미널(terminal) 집합
 - 예: $\{a, b\}$
- N: 논터미널(non-terminal) 집합
 - 예: $\{S, X, Y\}$
- P: 생성 규칙(production rule) 집합
 - 예: $\{S \rightarrow XY \mid X, X \rightarrow a \mid b, Y \rightarrow a \mid b\}$
- S: 시작 기호(start symbol)
 - 논터미널 기호 중 하나(첫번째 것)

5

정규 문법

- 모든 생성규칙의 LHS(Left-Hand Side)는 논터미널 1개
- 모든 생성규칙의 RHS(Right-Hand Side)는 터미널 스트링으로만 구성되거나 또는 논터미널 1개가 터미널 스트링 끝에 오른쪽 (혹은 왼쪽) 끝에 올 수 있음

6

정규 문법의 두 가지 유형

- $t_s = t_1 t_2 \dots t_n$, $n \geq 0$ 인 터미널 스트링이라 할 때
- 우선형 문법(Right-Linear Grammar)

$$\begin{aligned} A &\rightarrow t_s B \\ A &\rightarrow t_s \end{aligned}$$

- 좌선형 문법(Left-Linear Grammar)

$$\begin{aligned} A &\rightarrow B t_s \\ A &\rightarrow t_s \end{aligned}$$

- 예1, 예2, 예3은 정규 문법
- 예4, 예5는 정규 문법이 아님

7

정규 문법의 정의를 단순화

- 터미널 스트링을 터미널 1개로 제한

$$\text{우선형 문법: } A \rightarrow aB \mid b$$

$$\text{좌선형 문법: } A \rightarrow Ba \mid b$$

- $A \rightarrow t_1 t_2 \dots t_n B$ 유형은 아래와 같이 변환

$$\begin{aligned} A &\rightarrow t_1 A^1 \\ A^1 &\rightarrow t_2 A^2 \\ A^2 &\rightarrow t_3 A^3 \\ A^{n-1} &\rightarrow t_n B \end{aligned}$$

8

4가지 각각의 차이점 알기

- 문맥자유 문법(CFG) 왼쪽만 제약있음
 - 모든 생성규칙이 $A \rightarrow \alpha$ 유형
 - α 는 터미널 또는 논터미널로 구성된 스트링, $|\alpha| \geq 0$
- 문맥의존 문법(CSG) 더 자유로워짐 대신 크기 제약이 있음
 - 모든 생성규칙이 $\alpha \rightarrow \beta$ 유형이고, $|\alpha| \leq |\beta|$
 - α, β 는 터미널 또는 논터미널로 구성된 스트링, $|\alpha| \geq 1$
- 무제한 문법(UG) 아무 제약이 없음
 - 모든 생성규칙이 $\alpha \rightarrow \beta$ 유형
 - α, β 는 터미널 또는 논터미널로 구성된 스트링, $|\alpha| \geq 1$

9

문법 기술 방법

- 유한 언어: 스트링 개수가 유한 개
 - 모든 스트링을 나열하는 것이 가능함
 - 예1, 예2
- 무한 언어: 스트링 개수가 무한 개
 - 모든 스트링을 나열할 수 없음
 - 순환 규칙(recursive rule)으로 기술
 - 예3

10

1. (정규문법) $A \rightarrow aB : b$
2. (CFG) $A \rightarrow \alpha \mid \alpha \mid \dots \mid \alpha$
3. (CSG) 1. $\alpha \rightarrow \beta, \mid \alpha \mid \leq \mid \beta \mid$
4. (UG) $\alpha \rightarrow \beta$

문법 기술할 때 주의 사항

- 완전성(completeness)
 - 언어에 속하는 "모든 스트링을 생성"할 수 있어야 함
 - 즉, 1개라도 생성하지 못하는 스트링이 있으면 안됨
 - 건전성(soundness)
 - "그 언어에 속하는 스트링만 생성"해야 함
 - 즉, 언어에 속하지 않은 스트링을 1개라도 생성할 수 있으면 안됨
- $a = b;$
 $a = +++b;$

11

순환 규칙(recursive rule)

- 우순환 규칙(right recursive rule)

$$\begin{aligned} A &\rightarrow aA \mid b & A \Rightarrow aA \Rightarrow aaA \Rightarrow aaaA \Rightarrow \dots a \dots aA \\ & & A \text{에 } b \text{ 가능} \\ a^*b &= \{ b, a^1b, a^2b, \dots, a^nb, \dots \} \end{aligned}$$

- 좌순환 규칙(left recursive rule)

$$\begin{aligned} A &\rightarrow Aa \mid b & A \Rightarrow Aa \Rightarrow Aaa \Rightarrow Aa \dots a \\ ba^* &= \{ b, ba^1, ba^2, \dots, ba^n, \dots \} \end{aligned}$$

12

30가지 유형 문장
 29개만 있으면 안됨 (누락된게 있으면 안됨)
 그래서 하나 더추가했는데 모르고 또추가함
 29 + 1 + 1 이것도 안됨. 더도말고 부족해도안되고
 넘어가도안되고 완벽하게 딱 맞게

- 순환 규칙의 예

$$1) A \rightarrow aA \mid \varepsilon \quad \text{우}$$

$$2) A \rightarrow Aa \mid \varepsilon \quad \text{좌}$$

- 언어 $L_4 = \{ a^n b^n \mid n \geq 1 \}$ 에 대한 순환 규칙

$$A \rightarrow aAb \mid ab$$

13

문법 기술 연습

- $L_5 = \{ a^l b^m c^n \mid l, m, n \geq 1 \}$ 에 대한 정규 문법은?

- $L_6 = \{ a^l b^m c^n \mid l, m, n \geq 0 \}$ 에 대한 정규 문법은?

14

문법 기술 연습

- 덧셈 언어 $L_5 = \{ a, a+a, a+a+a, \dots \}$

$$A \rightarrow a+A \mid a$$

terminal들은 'a' '+'
쓰기도 함 명확하게하려면

- 문장 언어 $L_6 = \{ s; , s;s; , s;s;s; , \dots \}$

$$A \rightarrow s;A \mid s;$$

15

문법 기술 연습

- 괄호 언어

$$L_7 = \{ (a), ((a)), \dots, (a)(a), (a)((a)), \dots, ((a)), ((a))(a), \dots \}$$

$$A \rightarrow AA \mid (A) \mid (a)$$

AA = 열고닫고 자유자재 가능 $((()))(?)$

- 사칙연산 수식

$$E \rightarrow E + E \mid E - E \mid E * E \mid E / E \mid a$$

16

문법 기술 연습

- C 언어의 복문 $\{ s; s; \dots; s; \}$ 을 기술하는 문법은?

$$B \rightarrow \{ A \}$$

$$A \rightarrow s;A \mid s;$$

- 중첩된 if문을 기술하는 문법은?
- 조건식을 e, 문장을 s라고 함

$$C \rightarrow \text{if } (e) S \mid \text{if } (e) S \text{ else } S$$

$$S \rightarrow C \mid B \mid s;$$

17

유도(derivation)

- 유도(derivation)

- 시작기호로부터 터미널 스트링 생성 과정

- 예제

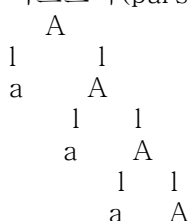
$$A \rightarrow aA \mid b$$

$$A \Rightarrow aA \Rightarrow aaA \Rightarrow aaaA \Rightarrow aaab$$

유도과정 => 을 사용

18

파스트리(parse tree)를 이용할 수 있음



문법 예제

- $G_1 = (\{O, E\}, \{a, b\}, P, O)$
 $P: \begin{array}{l} O \rightarrow a \mid bE \\ E \rightarrow aO \end{array}$
 길이가 짧은거부터 긴순 recursive하게
 $O \rightarrow a:baO$ (ba가 반복)
 $a, (ba)^n a, n \geq 0$
- $G_2 = (\{S, A, B, C\}, \{a, b, c, d\}, P, S)$
 $P: \begin{array}{l} S \rightarrow aA \\ A \rightarrow aA \mid bB \\ B \rightarrow bB \mid cC \\ C \rightarrow cC \mid d \end{array}$

19

문법 예제

- $G_3 = (\{S, A\}, \{a, b, c\}, P, S)$
 $P: \begin{array}{l} S \rightarrow bA \mid c \\ A \rightarrow aA \mid b \end{array}$
- $G_4 = (\{S, A\}, \{a, b, c\}, P, S)$
 $P: \begin{array}{l} S \rightarrow aSbA \mid c \\ A \rightarrow aAa \mid b \end{array}$
- $G_5 = (\{S, A\}, \{a, b\}, P, S)$
 $P: \begin{array}{l} S \rightarrow aAb \\ A \rightarrow aS \\ A \rightarrow b \end{array}$

20

형식 언어 예제

- $L = \{a^m b^n \mid m, n \geq 0\} : RG$
- $L_m = \{a^n b^n \mid n \geq 0\} : CFG$
- $L_{dm} = \{a^n b^n c^n \mid n \geq 0\} : CSG$
- $L_{mi} = \{\omega \omega^R \mid \omega \in V_T^*\} : \text{거울 언어(mirror language)}$
- $L_r = \{\omega \mid \omega = \omega^R\} : \text{회문 언어(palindrome)}$
- $L_p = \{\omega \mid \omega \text{는 balanced parenthesis로 구성}\} : \text{괄호 언어}$

21

둘의 차이점?
 palindrome은 abcba or abba
 거울언어 abba만 (abcba는 아님)