

오토마타 과제2 - 탑-다운 파싱

2022-18758 이재현

1. 프로그램 개요

형식 문법에 기반한 파싱 프로그램을 구현한다. 문법은 탑다운 파싱 방식을 사용한다. 파싱 단계에서 입력이 accept될 때까지의 과정을 보여준다. 주어진 문법을 바탕으로 파싱 테이블을 구성했고, 파서를 구현하였다. 파서는 입력 문자열을 읽고 스택을 사용해 파싱 규칙을 적용한다.

2. 프로그램 구조

- 파싱 테이블 - 각 변수가 알파벳을 만났을 때 어떤 규칙을 적용할지 결정함. Python dictionary 자료구조로 구현함.
- 파싱함수 - 입력 문자열을 스택 기반으로 파싱. 파생과정을 저장하여 출력하고 reject될 때는 "reject"를 출력.

3. 상세 설명

$$E \rightarrow E + T \mid E - T \mid T$$
$$T \rightarrow T * F \mid T / F \mid F$$
$$F \rightarrow (E) \mid A$$
$$A \rightarrow a \mid b \mid c \mid d \mid x \mid y \mid z \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$$

위 원래 문법에 대하여, 왼쪽에서 재귀가 반복되는 현상을 막기 위해 문법을 변형하였다.

$$E \rightarrow TB$$
$$B \rightarrow +TB \mid -TB \mid \epsilon$$
$$T \rightarrow FC$$
$$C \rightarrow *FC \mid /FC \mid \epsilon$$
$$F \rightarrow (E) \mid A$$
$$A \rightarrow a \mid b \mid c \mid d \mid x \mid y \mid z \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$$

추가된 변수 B와 C는 각각 +, - | *, / 기호를 파싱한다.

변형된 문법에 따라 파싱 테이블을 작성한다.

	(a..z,1...9	+,	*,/)	#
E	E->TB	E->TB				
B			B->+TB B->-TB	B->e	B->e	B->e
T	T->FC	T->FC				
C			C->e	C->*FC C->/FC	C->e	C->e
F	F->(E)	F->A				
A		A->a,b,c....				

B와 C의 경우 각각 대응되지 않는 연산자에 대해서는 입실론으로 파싱한다.

다음은 푸시다운 오토마톤의 작동 방식에 대해 설명한다.

1. 초기상태 - 스택에 시작 변수와 끝 기호가 포함된다.
2. 파싱 과정 - 스택의 상단 기호가 알파벳이면 입력 문자열의 현재 기호와 비교해 일치할 경우 스택에서 제거하고 입력 포인터를 이동시킨다. 상단이 변수이면 파싱 테이블을 참조해 입력 기호에 맞는 규칙을 적용하고 스택에 푸시한다. 입실론(e) 로 파생될 경우에 스택에는 추가하지 않고 그냥 제거한다.
3. 종료 조건 - 입력 문자열이 끝나고(#) 스택이 비면 accept된 것으로, 파싱 과정을 출력하고 종료한다. 반복문을 도는 중에 적용할 수 있는 규칙이 없을 경우 reject로 종료한다.