(Final exam score) ≠ $\frac{25}{40}$

replace the lower of your scores on exam 1 and 2. if it is higher.

parse Trees

↱ Last time it was mentioned that. Each nonterminal of a grammar represents a set of

$$\boxed{\text{grammer = context - free grammer = BNF specification. (in this course)}}$$

finite sequences of terminals.

The set of sequences represented by the starting nonterminal is called the language of / generated by the grammar. [This is regarded as the set of sequences that is defined by the grammer]

① ufpl::= ip·f

② ip::= d|ip d . $\quad$ "|" means "the left side is the same as in the previous production"
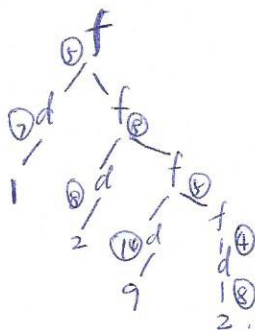
f::= d | df

d::= 0/1/2/3/ ... /9

The set of sequence of terminals that is represented by a nonterminal N is the set of all sequence of terminals that have parse trees rooted at N.

important

If the root of a parse tree is not specified. then it should be assumed that the root is the starting nonterminal.

Here is a parse tree that proves:

1 2 9 2 6 f

A parse tree rooted at a nonterminal N is an ordered rooted tree with the following properties:

1. The root is N

2. each leaf is either a terminal or the symbol <empty>; <empty> leaf has no siblings.

3. each internal node (i.e. each internal node) is a nonterminal. non-leaf.
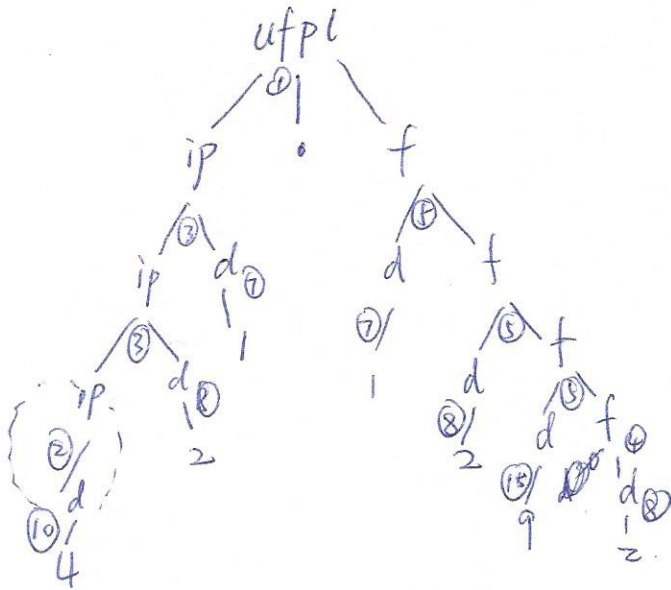
4. the sequence of children of each nonleaf node is the <u>right side of a production whose</u> <u>left side is that node</u>.

A parse tree whose sequence of non-<empty> leaves is $t_1, t_2 ... t_k$ is called a <u>parse tree</u> of or a parse tree that generates $t_1, t_2 ... t_k$.

Here is a parse tree that parses

$$4\ 2\ 1.\ 1\ 2\ 9\ 2 \in ufpl.$$

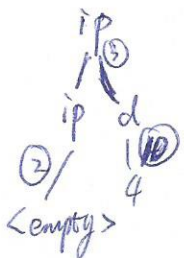[ie. $4\ 2\ 1.\ 1\ 2\ 9\ 2 \in$ the language of the above grammar].

Suppose we change production ② to $ip ::= \langle empty \rangle$

<u>Q</u>. Is the above tree still a correct parse tree? <u>No</u>

A correct parse tree of $4\ 2\ 1.\ 1\ 2\ 9\ 2$ should have the following subtree istead of

CS316 ③ 4/7/2016

Extended BNF (EBNF)

$EBNF = BNF + (\ldots)^+$ — that mean parentheses

$[\ldots]^+$ brackets — means "is optional"

$\{\ldots\}$ braces — "0 or more times"

brackets, and braces parentheses, in BNF has no special meaning; they are just ordinary ~~characters~~ [characters]

Ex: $e ::= (+|-)^+ (+|-)^+.$ is a EBNF rule that is equivalent to the following 4 BNF productions

$e ::= +t+t \mid -t+t \mid +t-t \mid -t-t$

$e ::= [+|-]t [(+|-)t]$ is equivalent to 9 BNF productions: $[\alpha] = \alpha | <empty>$

$e ::= t \mid t+t \mid t-t \mid +t \mid +t+t \mid +t-t \mid -t \mid -t+t \mid -t-t$

$e ::= [+|-]t \{(+|-)t\}$ is equivalent to infinitely many production, including:

$e ::= t+t+t-t-t-t+t.$
$e ::= -t-t+t+t+t+t-t-t$
$e ::= t.$

EBNF specifications can always be rewritten as a **finite** set of BNF productions.

One way to do this is to replace each occurrence of $(\ldots), [\ldots],$ and $\{\ldots\}$ with a new nonterminal, which you must define.

$(\ldots) \Rightarrow N$ where $N ::= \ldots$

$[\ldots] \Rightarrow N$ where $N ::= \ldots | <empty>$

$\{\ldots\} \Rightarrow N$ where $N ::= <empty> | N \ldots$

$\{\alpha | \beta | \gamma\} \Rightarrow <empty> | N(\alpha|\beta|\gamma) \Rightarrow N ::= <Empty> | N\alpha | N\beta | N\gamma.$

Work from the inside outwards

$e ::= \underbrace{[+|-]}_{opt\text{-}sign} t \underbrace{\{\overbrace{(+|-)}^{op}t\}}_{rest} \Rightarrow e ::= opt\text{-}sign\ t\ rest$

$opt\text{-}sign ::= +|-|<empty>$

$op ::= +|-$ $rest ::= <empty> | rest\ op\ t.$

Useful Equivalences

$$A ::= A\alpha \mid \beta \equiv A ::= \beta\{\alpha\}$$

$$A ::= \alpha A \mid \beta \equiv A ::= \{\alpha\}\beta$$

$$= \beta\alpha \dots \alpha$$
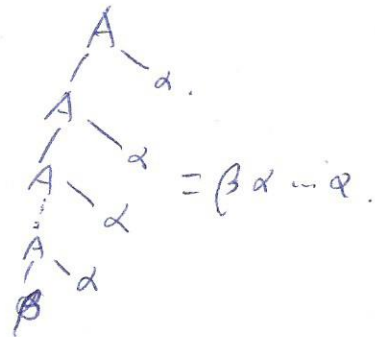
Ex: A seq of one or more <stmt>s in which consecutive <stmt>s are separated by a semicolon.

EBNF

$$\underset{A}{<\text{stmt-seq}>} ::= \underset{\beta}{<\text{stmt}>} \{\underset{\alpha}{; <\text{stmt}>}\} \qquad (\cancel{BNF}) \text{ BNF}: A ::= A\alpha \mid \beta.$$

$$\overset{A}{<\text{stmt-seq}>} ::= \overset{A}{<\text{stmt-seq}>}; \overset{\alpha}{\overbrace{<\text{stmt}>}} \mid <\text{stmt}>$$

seq of one or more d's

$$\overset{A}{ip} ::= d \underset{\alpha}{\underset{\beta}{\{d\}}} \Rightarrow ip ::= ip\,d \mid d.$$

$$\overset{A}{\cancel{f ::= \underset{\alpha}{\{d\}}\underset{\beta}{d}}} \Rightarrow \overset{A}{\cancel{f ::= d \mid f}}$$

$$\underset{A}{f} ::= \underset{\alpha}{\{d\}}\underset{\beta}{d} \Rightarrow f ::= d\,f \mid d.$$

$$\overset{A}{e} ::= \underset{\beta}{\overbrace{[+\mid-]}} t \underset{\alpha}{\overbrace{\{(+\mid-)\,t\}}}$$

$$\Rightarrow e ::= e \underbrace{(+\mid-)}\, t \mid [+\mid-]\,t$$

$$\Rightarrow e ::= e + t \mid e - t \mid + t \mid - t \mid t$$