<u>HackingOff</u>

- <u>Home</u>
- <u>Blog</u>
- Compiler Construction Toolkit
 - Overview
 - <u>Scanner Generator</u>
 - o Regex to NFA & DFA
 - NFA to DFA
 - o BNF to First, Follow, & Predict sets
 - 0

0

- <u>Parser Generator Overview</u>
- <u>LL(1) Parser Generator</u>
- LR(0) Parser Generator
- SLR(1) Parser Generator

Generate Predict, First, and Follow Sets from EBNF (Extended Backus Naur Form) Grammar

Provide a grammar in Extended Backus-Naur form (EBNF) to automatically calculate its first, follow, and predict sets. See the sidebar for an example.

First sets are used in LL parsers (top-down parsers reading Left-to-right, using Leftmost-derivations).

Follow sets are used in top-down parsers, but also in LR parsers (bottom-up parsers, reading \underline{L} eft-to-right, using \underline{R} ightmost derivations). These include LR(0), SLR(1), LR(k), and LALR parsers.

Predict sets, derived from the above two, are used by Fischer & LeBlanc to construct LL(1) top-down parsers.

Input Your Grammar

For more details, and a well-formed example, check out the sidebar. \rightarrow

```
compound-statement
if-statement | while-
statement | break-
statement
            continue
statement
            return-
statement
expression-statement
|declaration-
statement
if\text{-statement} \,\to\, if
expression compound-
statement else
compound-statement
while-statement -
while expression
compound-statement
break-statement ->
compound-statement ->
Click for Predict, First, and Follow Sets
```

First Set

```
First Set
    Non-Terminal Symbol
if
                                 if
                                 else
else
                                 while
while
                                 break
break
                                 }
{\tt continue}
                                 continue
return
                                 return
*=
                                 *=
                                  /=
/=
+=
                                 +=
-=
                                 -=
                                 &&=
&&=
хх=
                                 ХХ=
```

```
&&
&&
                              ==
==
!=
                              !=
                              <
<
<=
                              <=
                              ++
identifier
                              identifier
(
                              (
INT-LITERAL
                              INT-LITERAL
                              BOOL-LITERAL
BOOL-LITERAL
                              var
class
                              class
const
                              const
int
                              int
                              boo1
boo1
                              if
if-statement
                              while
while-statement
                              break
break-statement
compound-statement
                              ε, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool
statement-list
continue-statement
                              continue
return-statement
expression-statement
                              ;, ε, -, !, ++, --
expression-list
                              ε, -, !, ++, --
class-body
variable-declaration-list
                              ε, var, int, bool
                              =, *=, /=, +=, -=, &&=, XX=
assignment-operator
condition-or-expression-tail \ \epsilon , XX
condition-and-expression-tail &&, ε
                             ε, ==, !=
equality-expression-tail
                              ε, <, <=, >, >=
rel-expression-tail
additive-expression-tail
                              ε, +, -
                              ε, *, /
m-d-expression-tail
                              -, !, ++, --
u-expression
post-expression-tail
                              ., ++
                              identifier, (, INT-LITERAL, BOOL-LITERAL
primary-expression
para-list
                              (
proper-para-list-tail
                              (
arg-list
proper-arg-list-tail
                              ,, ε
                              identifier
function-declaration
variable-declaration
                              var, int, bool
class-declaration
                              class
constant-declaration
                              const
init-expression
type-annotation
type
                              int, bool
                              \epsilon, \{, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool
top-level
                              \{, while, continue, if, return, break, ;, \epsilon, -, !, ++, --, identifier, const, class, var, int, bool
statement
m-d-expression \\
                              -, !, ++, --
                              identifier, (, INT-LITERAL, BOOL-LITERAL
post-expression
para-declaration
                              int, bool
declaration-statement
                              identifier, const, class, var, int, bool
additive-expression
                              -, !, ++,
proper-para-list
                              int, bool
rel-expression
                              -, !, ++, --
equality-expression
                              -, !, ++, --
\verb|condition-and-expression||
                              -, !, ++, --
{\tt condition-or-expression}
                              -, !, ++, --
```

```
assignment-expression -, !, ++, --
expression -, !, ++, --
arg -, !, ++, --
proper-arg-list -, !, ++, --
```

Follow Set

```
Follow Set
   Non-Terminal Symbol
                            $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
statement
                            $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
if-statement
                            $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
while-statement
                            $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
break-statement
                            else, $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int,
compound-statement
                            bool.
statement-list
                            $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
continue-statement
                            , {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
return-statement
                            $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
expression-statement
expression-list
class-body
variable-declaration-list
                            ), ;, {, ,
expression
                            ), ;, {, ,
assignment-expression\\
assignment-operator
condition-or-expression
                            ), ;, {, ,
condition-or-expression-
                            ), ;, {, ,
tail
                            XX, ), ;, {, ,
condition-and-expression
condition-and-expression-
                            XX, ), ;, {, ,}
tail
                            ==, !=, &&, XX, ), ;, {, ,
equality-expression
                            ==, !=, &&, XX, ), ;, {, ,
equality-expression-tail
                            ==, !=, &&, XX, ), ;, {, ,
rel-expression
                            ==, !=, &&, XX, ), ;, {, ,
rel-expression-tail
additive-expression
                            <, <=, >, >=, ==, !=, &&, XX, ), ;, {, ,
                            <, <=, >, >=, ==, !=, &&, XX, ), ;, {, ,
additive-expression-tail
                            +, -, <, <=, >, >=, ==, !=, &&, XX, ), ;, {, ,
m-d-expression
                            +, -, <, <=, >=, ==, !=, &&, XX, ), ;, {, ,
m-d-expression-tail
u-expression
                            *, /, +, -, <, <=, >, >=, ==, !=, &&, XX, ), ;, {, ,
post-expression
post-expression-tail
primary-expression
                            {
para-list
                            )
proper-para-list
proper-para-list-tail
para-declaration
                               )
arg-list
proper-arg-list
                            )
proper-arg-list-tail
arg
declaration-statement
                            $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
function-declaration
                            $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
variable-declaration
                            var, int, bool, $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, }
class-declaration
                            $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
                            $, {, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool, }
constant-declaration
init-expression
type-annotation
                            identifier, ;
type
top-level
```

Predict Set

```
Predict
                                     Expression
                                                                                     {
1 statement → compound-statement
                                                                                    if
2 statement \rightarrow if-statement
3 statement \rightarrow while-statement
                                                                                    while
   statement → break-statement
4
                                                                                    break
5
    \mathtt{statement} \, \rightarrow \, \mathtt{continue}\text{-}\mathtt{statement}
                                                                                    continue
    statement \rightarrow return-statement
                                                                                    return
```

```
statement → expression-statement
                                                                         :. -. !. ++. --
   statement \rightarrow declaration-statement
                                                                         identifier, const, class, var, int, bool
8
   if\text{--statement} \, \rightarrow \, if \, expression \, compound\text{--statement} \, \, else
    compound-statement
10 while-statement → while expression compound-statement
                                                                         while
11 break-statement → break;
                                                                         break
12 compound-statement → { statement-list }
                                                                         {
13 statement-list → ε
                                                                         {, while, continue, if, return, break, ;, -, !, ++, --,
14 statement-list → statement statement-list
                                                                         identifier, const, class, var, int, bool
15 continue-statement \rightarrow continue;
                                                                         continue
16 return-statement \rightarrow return expression;
                                                                         return
17 return-statement \rightarrow return ;
                                                                         return
18 expression-statement \rightarrow expression-list;
                                                                         -, !, ++, --, ;
19 expression—list → expression
                                                                         -, !, ++, --
20 expression-list \rightarrow \epsilon
21 class-body → { variable-declaration-list }
   variable-declaration-list → variable-declaration variable-
                                                                         var, int, bool
   declaration-list
23 variable-declaration-list → ε
24 expression → assignment-expression
                                                                         -, !, ++, --
25 assignment-expression → condition-or-expression
                                                                         -, !, ++, -
26 assignment-operator → =
27 assignment-operator → *=
28 assignment-operator → /=
                                                                         /=
29 assignment-operator → +=
                                                                         +=
30 assignment-operator → -=
                                                                         -=
31 assignment-operator → &&=
                                                                         &&=
32 assignment-operator \rightarrow XX=
                                                                         XX =
   {\tt condition-or-expression} \,\, {\color{red} \rightarrow} \,\, {\tt condition-and-expression}
                                                                         -, !, ++, --
    condition-or-expression-tail
34 condition-or-expression-tail → ε
                                                                         ), ;, {, ,
   condition-or-expression-tail \rightarrow XX condition-and-expression
35
   condition-or-expression-tail
   condition-and-expression → equality-expression condition-
                                                                         -, !, ++, --
   and-expression-tail
   condition—and—expression—tail → && equality—expression
                                                                         &&
    equality-expression-tail
   condition-and-expression-tail \rightarrow \epsilon
                                                                         XX, ), ;, {, ,
   equality-expression \rightarrow rel-expression equality-expression-
39
                                                                         -, !, ++, --
40 equality-expression-tail → ε
                                                                         ==, !=, &&, XX, ), ;, {, ,
   equality-expression-tail \rightarrow == rel-expression equality-
    expression-tail
   equality-expression-tail → != rel-expression equality-
   expression-tail
                                                                         -, !, ++, --
43 rel-expression → additive-expression rel-expression-tail
                                                                         ==, !=, &&, XX, ), ;, {, ,
44 rel-expression-tail → ε
   rel-expression-tail \rightarrow < additive-expression rel-expression-
45
    tail
   rel-expression-tail → <= additive-expression rel-expression-
46
   rel-expression-tail \rightarrow > additive-expression rel-expression-
47
   tail
   \texttt{rel-expression-tail} \; \rightarrow \; \texttt{>=} \; \texttt{additive-expression} \; \; \texttt{rel-expression-}_{\texttt{>=}}
   tail
   additive-expression \  \, \textbf{-} \  \, \textbf{m-}d-expression \  \, additive-expression-
49
                                                                         -, !, ++, --
    tail
50 additive-expression-tail \rightarrow \epsilon
                                                                         <, <=, >, >=, ==, !=, &&, XX, ), ;, {, ,
   additive-expression-tail \ \rightarrow \ + \ m-d-expression \ additive-
    expression-tail
   additive-expression-tail \rightarrow - m-d-expression additive-
   expression-tail
53 m-d-expression → u-expression m-d-expression-tail
                                                                         -, !, ++, --
                                                                         +, -, <, <=, >, >=, ==, !=, &&, XX, ), ;, {, ,
54 m-d-expression-tail → ε
55 m-d-expression-tail \rightarrow * u-expression m-d-expression-tail
56 m-d-expression-tail \rightarrow / u-expression m-d-expression-tail
57 u-expression \rightarrow - u-expression
58 u-expression \rightarrow ! u-expression
59 u-expression \rightarrow ++ u-expression
60 u-expression \rightarrow -- u-expression
61 post-expression → primary-expression
                                                                         identifier, (, INT-LITERAL, BOOL-LITERAL
62 post-expression \rightarrow primary-expression post-expression-tail
                                                                         identifier, (, INT-LITERAL, BOOL-LITERAL
```

```
63 post-expression-tail \rightarrow . identifier post-expression-tail
64 post-expression-tail → ++ post-expression-tail
65 primary-expression → identifier
                                                                       identifier
66 primary-expression → identifier arg-list
                                                                       identifier
67 primary-expression \rightarrow ( expression )
68 primary-expression \rightarrow INT-LITERAL
                                                                       INT-LITERAL
69 primary-expression → BOOL-LITERAL
                                                                       BOOL-LITERAL
70 para-list \rightarrow ()
71 para-list \rightarrow ( proper-para-list )
                                                                       (
72 proper-para-list → para-declaration proper-para-list-tail
                                                                       int. bool
   proper-para-list-tail → , para-declaration proper-para-list-
73
   tail
74 proper-para-list-tail → ε
75 para-declaration → type identifier
                                                                       int, bool
76 arg-list \rightarrow ()
                                                                       (
77 arg-list → ( proper-arg-list )
78 proper-arg-list \rightarrow arg proper-arg-list-tail
79 proper-arg-list-tail \rightarrow , arg proper-arg-list-tail
80 proper-arg-list-tail \rightarrow \epsilon
                                                                       )
81 arg → expression
                                                                       -, !, ++, --
82 declaration-statement → function-declaration
                                                                       identifier
83 declaration-statement \rightarrow constant-declaration
                                                                       const
84 declaration-statement → variable-declaration
                                                                       var, int, bool
85 declaration-statement → class-declaration
                                                                       class
   function-declaration → identifier para-list compound-
86
                                                                       identifier
   statement
87 variable-declaration \rightarrow var identifier init-expression ;
                                                                       var
88 variable-declaration \rightarrow var identifier type-annotation ;
                                                                       var
89 variable-declaration \rightarrow type identifier init-expression ;
                                                                       int, bool
90 variable-declaration \rightarrow type identifier;
                                                                       int, bool
91 class-declaration \rightarrow class identifier init-expression;
                                                                       class
92 class-declaration \rightarrow class identifier type-annotation;
                                                                       class
93 constant-declaration \rightarrow const identifier init-expression;
                                                                       const
94 constant-declaration \rightarrow const identifier type-annotation;
95 init-expression \rightarrow = expression
96 type-annotation \rightarrow : type
97
   type → int
                                                                       int
98 type → bool
                                                                       hoo1
                                                                       \{, while, continue, if, return, break, ;, -, !, ++, --,
99 top-level → statement top-level
                                                                       identifier, const, class, var, int, bool
100 top-level → \epsilon
```

LL(1) Parsing Table

On the LL(1) Parsing Table's Meaning and Construction

- The top row corresponds to the columns for all the potential terminal symbols, augmented with \$ to represent the end of the parse.
- The leftmost column and second row are all zero filled, to accommodate the way Fischer and LeBlanc wrote their parser's handling of abs().
- The remaining rows correspond to production rules in the original grammar that you typed in.
- Each entry in that row maps the left-hand-side (LHS) of a production rule onto a line-number. That number is the line in which the LHS had that specific column symbol in its predict set.
- If a terminal is absent from a non-terminal's predict set, an error code is placed in the table. If that terminal is in follow(that non-terminal), the error is a POP error. Else, it's a SCAN error.

```
POP error code = # of predict table productions + 1
SCAN error code = # of predict table productions + 2
```

In practice, you'd want to tear the top, label row off of the table and stick it in a comment, so that you can make sense of your table. The remaining table can be used as is.

LL(1) Parsing Table as JSON (for Easy Import)

LL(1) Parsing Push-Map (as JSON)

This structure maps each production rule in the expanded grammar (seen as the middle column in the predict table above) to a series of states that the LL parser pushes onto the stack.

How to Calculate First, Follow, & Predict Sets

Specify your grammar in EBNF and slam the button. That's it.

EBNF Grammar Specification Requirements

Productions use the following format:

```
Goal -> A
A \rightarrow (A) \mid Two
Two -> a
Two -> b
```

- Symbols are inferred as terminal by absence from the left hand side of production rules. "->" designates definition, "|" designates alternation, and newlines designate termination.
- -> y | z is EBNF short-hand for x -> y
- Use "EPSILON" to represent ϵ or "LAMBDA" for λ productions. (The two function identically.) E.g., A -> b | EPSILON.
- Be certain to place spaces between things you don't want read as one symbol. (A) \neq (A)

About This Tool

Intended Audience

Computer science students & autodidacts studying compiler design or parsing.

Purpose

Automatic generation of first sets, follow sets, and predict sets speeds up the process of writing parsers. Generating these sets by hands is tedious; this tool helps ameliorate that. Goals:

- Tight feedback loops for faster learning.
 Convenient experimentation with language tweaks. (Write a generic, table/dictionary-driven parser and just plug in the JSON output to get off the ground quickly.)
- Help with tackling existing coursework or creating new course material.

Underlying Theory

 $\mbox{I'll}\ \mbox{do}\ \mbox{a}$ write-up on this soon. In the interim, you can read about:

- how to determine first and follow sets (PDF from Programming Languages course at University of Alaska Fairbanks)
- <u>significance</u> of first and follow sets in top-down (LL(1)) parsing.
- follow sets' involvement in bottom-up parsing (LALR, in this case)
- © HackingOff.com 2012