

[HackingOff](#)

- [Home](#)
- [Blog](#)
- [Compiler Construction Toolkit](#)
  - [Overview](#)
  - [Scanner Generator](#)
  - [Regex to NFA & DFA](#)
  - [NFA to DFA](#)
  - [BNF to First, Follow, & Predict sets](#)
  - [Parser Generator Overview](#)
  - [LL\(1\) Parser Generator](#)
  - [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 Left-to-right, using Rightmost 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. →

statement ->  
compound-statement |  
if-statement | while-  
statement | break-  
statement | continue-  
statement | return-  
statement |  
expression-statement  
| declaration-  
statement  
if-statement -> if  
expression compound-  
statement else  
compound-statement  
while-statement ->  
while expression  
compound-statement  
break-statement ->  
break ;  
compound-statement ->

Click for Predict, First, and Follow Sets

## First Set

Non-Terminal	Symbol	First Set
if	if	
else	else	
while	while	
break	break	
;	;	
{	{	
}	}	
ε	ε	
continue	continue	
return	return	
=	=	
*=	*=	
/=	/=	
+=	+=	
-=	-=	
&&=	&&=	
XX=	XX=	
XX	XX	

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

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

## Follow Set

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

## Predict Set

#	Expression	Predict
1	statement → compound-statement	{
2	statement → if-statement	if
3	statement → while-statement	while
4	statement → break-statement	break
5	statement → continue-statement	continue
6	statement → return-statement	return

7	statement → expression-statement	;, -, !, ++, --
8	statement → declaration-statement	identifier, const, class, var, int, bool
9	if-statement → if expression compound-statement else compound-statement	if
10	while-statement → while expression compound-statement	while
11	break-statement → break ;	break
12	compound-statement → { statement-list }	{
13	statement-list → ε	}
14	statement-list → statement statement-list	{, while, continue, if, return, break, ;, -, !, ++, --, identifier, const, class, var, int, bool
15	continue-statement → continue ;	continue
16	return-statement → return expression ;	return
17	return-statement → return ;	return
18	expression-statement → expression-list ;	-, !, ++, --, ;
19	expression-list → expression	-, !, ++, --
20	expression-list → ε	;
21	class-body → { variable-declaration-list }	{
22	variable-declaration-list → variable-declaration variable-declaration-list	var, int, bool
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 → XX=	XX=
33	condition-or-expression → condition-and-expression condition-or-expression-tail	-, !, ++, --
34	condition-or-expression-tail → ε	), ;, {, ,
35	condition-or-expression-tail → XX condition-and-expression condition-or-expression-tail	XX
36	condition-and-expression → equality-expression condition-and-expression-tail	-, !, ++, --
37	condition-and-expression-tail → && equality-expression equality-expression-tail	&&
38	condition-and-expression-tail → ε	XX, ), ;, {, ,
39	equality-expression → rel-expression equality-expression-tail	-, !, ++, --
40	equality-expression-tail → ε	==, !=, &&, XX, ), ;, {, ,
41	equality-expression-tail → == rel-expression equality-expression-tail	==
42	equality-expression-tail → != rel-expression equality-expression-tail	!=
43	rel-expression → additive-expression rel-expression-tail	-, !, ++, --
44	rel-expression-tail → ε	==, !=, &&, XX, ), ;, {, ,
45	rel-expression-tail → < additive-expression rel-expression-tail	<
46	rel-expression-tail → <= additive-expression rel-expression-tail	<=
47	rel-expression-tail → > additive-expression rel-expression-tail	>
48	rel-expression-tail → >= additive-expression rel-expression-tail	>=
49	additive-expression → m-d-expression additive-expression-tail	-, !, ++, --
50	additive-expression-tail → ε	<, <=, >, >=, ==, !=, &&, XX, ), ;, {, ,
51	additive-expression-tail → + m-d-expression additive-expression-tail	+
52	additive-expression-tail → - m-d-expression additive-expression-tail	-
53	m-d-expression → u-expression m-d-expression-tail	-, !, ++, --
54	m-d-expression-tail → ε	+, -, <, <=, >, >=, ==, !=, &&, XX, ), ;, {, ,
55	m-d-expression-tail → * u-expression m-d-expression-tail	*
56	m-d-expression-tail → / u-expression m-d-expression-tail	/
57	u-expression → - u-expression	-
58	u-expression → ! u-expression	!
59	u-expression → ++ u-expression	++
60	u-expression → -- u-expression	--
61	post-expression → primary-expression	identifier, (, INT-LITERAL, BOOL-LITERAL
62	post-expression → primary-expression post-expression-tail	identifier, (, INT-LITERAL, BOOL-LITERAL



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

[1":[5], "2":[2], "3":[3], "4":[4], "5":[7], "6":[8], "7":[9], "8":[40], "9":[5,-2,5,13,-1], "10":[5,13,-3], "11":[-5,-4], "12":  
[-7,6,-6], "14":[6,1], "15":[-5,-8], "16":[-5,13,-9], "17":[-5,-9], "18":[-5,10], "19":[13], "21":[-7,12,-6], "22":[12,42], "24":[14], "25":  
[16], "26":[-10], "27":[-11], "28":[-12], "29":[-13], "30":[-14], "31":[-15], "32":[-16], "33":[17,18], "35":[17,18,-17], "36":[19,20], "37":  
[21,20,-18], "39":[21,22], "41":[21,22,-19], "42":[21,22,-20], "43":[23,24], "45":[23,24,-21], "46":[23,24,-22], "47":[23,24,-23], "48":  
[23,24,-24], "49":[25,26], "51":[25,26,-25], "52":[25,26,-26], "53":[27,28], "55":[27,28,-27], "56":[28,28,-28], "57":[28,-26], "58":  
[28,-29], "59":[28,-30], "60":[28,-31], "61":[31], "62":[30,31], "63":[30,-33,-32], "64":[30,-30], "65":[-33], "66":[36,-33], "67":  
[-35,13,-34], "68":[-36], "69":[-37], "70":[-35,-34], "71":[-35,33,-34], "72":[34,35], "73":[34,35,-38], "75":[-33,47], "76":[-35,-34], "77":  
[-35,37,-34], "78":[38,39], "79":[38,39,-38], "81":[13], "82":[41], "83":[44], "84":[42], "85":[43], "86":[5,32,-33], "87":  
[-5,45,-33,-39], "88":[-5,46,-33,-39], "89":[-5,45,-33,47], "90":[-5,-33,47], "91":[-5,45,-33,-40], "92":[-5,46,-33,-40], "93":  
[-5,45,-33,-41], "94":[-5,46,-33,-41], "95":[13,-10], "96":[47,-42], "97":[-43], "98":[-44], "99":[48,1]]

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

Productions use the following format:

- Symbols are inferred as terminal by absence from the left hand side of production rules.
- “ $\rightarrow$ ” designates definition, “ $|$ ” designates alternation, and newlines designate termination.
- $x \rightarrow y \mid z$  is EBNF short-hand for
 
$$\begin{array}{l} x \rightarrow y \\ x \rightarrow z \end{array}$$
- Use “EPSILON” to represent  $\epsilon$  or “LAMBDA” for  $\lambda$  productions. (The two function identically.) E.g.,  $A \rightarrow b \mid \text{EPSILON}$ .
- Be certain to place spaces between things you don’t want read as one symbol.  $(A) \neq (A)$

## 6/7

## 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

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

- [how to determine first and follow sets \(PDF from Programming Languages course at University of Alaska Fairbanks\)](#)
- [significance 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