Project 2: Grammar Analysis and Parsing

S. Patel, J. Collard, M. Barney April 18, 2013

1 Introduction

This report contains our implementation of a scanner and parser for context free grammars, a series of hygiene functions for sterilizing the grammar, and finally a parser *for* the grammar specified in the context free grammar.

It is divided up into several sections, roughly corresponding to the problems given in the specification, each a Haskell module. The work was split up evenly amongst the group members, and approximately 40 man hours went into the final preparation of this document, the source code, unit testing, and related work.

2 Context Free Grammar

In this section we provide the context free grammar data type.

At its heart, a grammar it consists of a list of productions, where each production consists of a constructor and two arguments; the first a paramaterized nonterminal, and the second a paramaterized right hand side.

An *RHS* is either empty, a terminal, which takes two arguments — the paramaterized object representing a terminal, and another *RHS*; or a non-terminal, which similarly takes two arguments.

```
{-# LANGUAGE FlexibleInstances, MultiParamTypeClasses #-} module ContextFreeGrammar (Grammar, Production (...), RHS (...), module Dropable, nonTerminals, terminals, Terminal (...)) where import <math>Dropable import Filterable import Filterable import Prelude\ hiding\ (drop\ filter) type Grammar\ nt\ t = [Production\ nt\ t]
```

```
data Terminal t = Epsilon \mid EOF \mid Terminal \ t \ deriving (Show, Eq. Ord)
      instance (Eq\ nt) \Rightarrow Dropable\ nt\ (Grammar\ nt\ t) where
         drop \ x \ grammar = map \ (drop \ x) \ grammar
      instance Filterable (nt \rightarrow Bool) (Grammar nt \ t) where
         filter\ pred\ grammar = map\ (filter\ pred)\ grammar
      data Production nt \ t = Production \{ nonterminal :: nt, \}
         rhs :: RHS \ nt \ t \} \ deriving (Eq, Ord, Show)
      instance (Eq\ nt) \Rightarrow Dropable\ nt\ (Production\ nt\ t) where
         drop \ x \ (Production \ nt \ rhs) = Production \ nt \ (drop \ x \ rhs)
      instance Filterable (nt \rightarrow Bool) (Production nt t) where
         filter\ pred\ (Production\ nt\ rhs) = Production\ nt\ (filter\ pred\ rhs)
      data RHS nt t = Empty
           Term t (RHS nt t)
         | NonT \ nt \ (RHS \ nt \ t) \ deriving \ (Eq, Ord, Show)
      instance (Eq \ nt) \Rightarrow Dropable \ nt \ (RHS \ nt \ t) where
         drop \ x \ (NonT \ nt \ rhs)
            | x \equiv nt = drop \ x \ rhs
            | otherwise = (NonT \ nt \ (drop \ x \ rhs))
         drop \ x \ (Term \ t \ rhs) = Term \ t \ (drop \ x \ rhs)
         drop \ \_Empty = Empty
      instance Filterable (nt \rightarrow Bool) (RHS nt\ t) where
        filter \_Empty = Empty
         filter\ pred\ (Term\ t\ rhs) = (Term\ t\ (filter\ pred\ rhs))
        filter\ pred\ (NonT\ nt\ rhs) =
           if pred nt then (NonT nt (filter pred rhs)) else (filter pred rhs)
nonTerminals takes the RHS of a Production and returns a list of all Non
Terminals
      nonTerminals :: RHS \ nt \ t \rightarrow [nt]
      nonTerminals (NonT nt rhs) = nt : nonTerminals rhs
      nonTerminals (Term \_ rhs) = nonTerminals rhs
      nonTerminals\ Empty = []
terminals takes the RHS of a Production and returns a list of all Terminals
      terminals :: RHS \ nt \ t \rightarrow [t]
      terminals (Term \ t \ rhs) = t : terminals \ rhs
      terminals (NonT - rhs) = terminals rhs
```

```
terminals \ Empty = []
simple Grammar :: Grammar \ String \ String
simple Grammar = [a, b, c, d] \ \mathbf{where}
a = Production \ "A" \ (Term \ "a" \ Empty)
b = Production \ "B" \ (NonT \ "B" \ Empty)
c = Production \ "C" \ (Term \ "a" \ (NonT \ "B" \ Empty))
d = Production \ "D" \ (NonT \ "B" \ (Term \ "a" \ Empty))
```

3 Scanner and Parser for context-free grammars

In this section we provide code for a simple scanner and parser for a textual representation of a context free grammar.

The grammar for the concrete representation follows the suggestion in the assignment, with one minor difference:

```
Grammar -> Grammar Production
Grammar -> Production
Production -> UpperSymbol Arrow RHS
RHS -> RHS Symbol
RHS ->
Symbol -> UpperSymbol
Symbol -> AlphaNumSymbol
```

In other words, non-terminals are restricted to their first letter being upper case, terminals are sequences of alphanumeric characters where the first character cannot be upper-case, and right hand side terminals and non-terminals are delimited by spaces.

A couple helper functions are initially defined, in addition to the grammar token data structure, which is as follows:

```
 \begin{array}{l} \textbf{module } ScanAndParse \ (sparse) \ \textbf{where} \\ \textbf{import } ContextFreeGrammar \\ \textbf{import } Data. Char \ (isUpper, isSpace, isAlphaNum, isAlpha, isDigit) \\ \textbf{data } GrammarToken = \\ Symbol \ String \mid \\ ArrowToken \mid \\ NewLineToken \ \textbf{deriving } \ (Show, Eq) \\ alphanumeric = takeWhile \ isAlphaNum \\ \end{array}
```

```
drop' _{-}[] = []
drop' i (x : xs) =
if i \leq 0 then (x : xs)
else
drop' (i - 1) xs
```

The scanner is a simple function that checks for two special characters, the arrow, \rightarrow and the newline character, n, scans symbols for nonterminals or terminals, and returns their appropriate tokens.

If a non alphanumeric character is found, the scanner returns an error.

```
 scan :: String \rightarrow [GrammarToken] \\ scan [] &= [] \\ scan ('-'::'>':cs) &= ArrowToken : scan \ cs \\ scan ('\setminus n':cs) &= NewLineToken : scan \ cs \\ scan (c:cs) \mid isSpace \ c = scan \ cs \\ scan s@(c:cs) \mid isAlphaNum \ c = \\ \textbf{let} \ name = alphanumeric \ s \\ len = length \ name \ \textbf{in} \\ (Symbol \ name) : scan \ (drop' \ len \ s) \\ scan s@(c:cs) &= \\ error ("lexical \ error; " ++ c:" \ \textbf{is} \ \textbf{an} \ unrecognized \ character.")
```

The parser generates a list of productions, i.e., a "grammar", from a list of grammar tokens. The helper function, *parseRHS*, will throw a syntax error if an arrow token is found on the right hand side.

The function *parse* will throw an error if multiple non-terminals occur on the left-hand side, or an arrow is missing.

```
parseRHS :: [GrammarToken] \rightarrow ((RHS\ String\ String), [GrammarToken])
parseRHS\ [] = (Empty, [])
parseRHS\ (NewLineToken: rhs) = (Empty, rhs)
parseRHS\ (ArrowToken: rhs) = error\ "syntax\ error;\ arrow\ token\ found\ on\ right\ hand\ side"
parseRHS\ ((Symbol\ (c:cs)): rhs) = 
let\ (term, rhs') = parseRHS\ rhs\ in
if\ isUpper\ c\ then
((NonT\ (c:cs)\ term), rhs')
else
```

```
((Term\ (c:cs)\ term), rhs')
parse :: [GrammarToken] \rightarrow Grammar\ String\ String
parse\ [] = []
parse\ (NewLineToken:p) = parse\ p
parse\ ((Symbol\ s): ArrowToken: rhs) =
let\ (production, rhs') = parseRHS\ rhs\ in
(Production\ s\ (production)): parse\ rhs'
parse\ ((Symbol\ s): rhs) =
error\ "Missing\ arrow\ or\ multiple\ non-terminals\ on\ left-hand\ side."
sparse = parse \circ scan
```

4 Hygiene Module

In this module, we perform basic hygiene checks on the grammar, remove unreachable non terminals, etc.

```
module BadHygiene (computeReachable,
eliminateUnreachable,
computeGenerating,
eliminateNonGenerating,
eliminateUseless,
isEmptyGrammar) where
import ContextFreeGrammar
import qualified Data.Set as S
import Filterable
import ScanAndParse
{-BEGIN CLEANING FUNCTIONS -}
```

computeReachable finds the Set of all Non Terminals of a Grammar that can be reached from the start node.

```
compute Reachable :: Ord \ nt \Rightarrow Grammar \ nt \ t \rightarrow S.Set \ nt compute Reachable \ [] = S.empty compute Reachable \ ps = go \ (S.singleton \circ nonterminal \circ head \ ps) \ (concat \circ replicate \ (length \ go \ marked \ [] = marked go \ marked \ ((Production \ nt \ rhs) : prs) = \mathbf{if} \ S.member \ nt \ marked \mathbf{then} \ go \ marked' \ prs \mathbf{else} \ go \ marked \ prs \mathbf{where} \ marked' = S.union \ marked \circ S.fromList \circ nonTerminals \ rhs
```

eliminate Unreachable removes all unreachable Non Terminals from a Grammar.

```
eliminate Unreachable :: Ord nt \Rightarrow Grammar \ nt \ t \rightarrow Grammar \ nt \ t
eliminate Unreachable g = clean Grammar \ where
reachable = compute Reachable \$ g
-- unnecessary? By definition, the unreachable non-terminals cannot be in any
-- other production list.
-- clean Productions = Filterable. filter ('S. member' reachable) g
clean Grammar = Prelude. filter (\lambda(Production \ nt \ rhs) \rightarrow S. member \ nt \ reachable) g
```

compute Generating finds the Set of all Non Terminals of a Grammar that can produce a string of Terminals.

```
 \begin{array}{l} compute Generating :: (Ord \ nt, Ord \ t) \Rightarrow Grammar \ nt \ t \rightarrow S.Set \ nt \\ compute Generating \ [] = S.empty \\ compute Generating \ ps = go \ S.empty \ (concat \circ replicate \ (length \ ps) \ ps) \ \textbf{where} \\ all Terms = S.from List \circ concat Map \ (terminals \circ rhs) \ ps \\ go \ marked NT \ [] = marked NT \\ go \ marked NT \ ((Production \ nt \ rhs) : prs) = \textbf{if} \ (all \ (`S.member`all Terms) \circ terminals \ rhs) \\ (all \ (`S.member`marked NT) \circ non Terminals \ rhs) \\ \textbf{then} \ go \ (S.insert \ nt \ marked NT) \ prs \\ \textbf{else} \ go \ marked NT \ prs \end{array}
```

eliminateNonGenerating removes all non Generating Non Terminals from a Grammar.

```
eliminateNonGenerating :: (Ord nt, Ord t) \Rightarrow Grammar nt t \rightarrow Grammar nt t
eliminateNonGenerating g = cleanGrammar where
generating = computeGenerating g
cleanProductions = Filterable.filter ('S.member'generating) g
cleanGrammar = Prelude.filter (\lambda(Production nt rhs) \rightarrow S.member nt generating) <math>cleanProduction
```

eliminate Useless removes all non Generating and unreachable Non Terminals from a Grammar.

```
eliminateUseless :: (Ord\ nt, Ord\ t) \Rightarrow Grammar\ nt\ t \rightarrow Grammar\ nt\ t
eliminateUseless = eliminateUnreachable \circ eliminateNonGenerating
```

is Empty Grammar determines if a Grammar will produce any strings at all.

```
isEmptyGrammar :: (Ord\ t, Ord\ nt) \Rightarrow Grammar\ nt\ t \rightarrow Bool isEmptyGrammar\ [] = True isEmptyGrammar\ g = \neg \circ elem\ nt \circ map\ nonterminal\ \g' where g' = eliminateNonGenerating\ g (Production\ nt\ \_) = head\ g {-END CLEANING FUNCTIONS -}
```

5 Nullable, First, and Follow

In this section, we provide several modules for computing the nullable, first and follow sets of a given context free grammar, respectively.

5.1 Nullable

Here we compute whether a production is nullable or not.

```
module Nullable (nullable) where
import ContextFreeGrammar
import qualified Data. Set as S
import Prelude hiding (drop)
type Set = S.Set
nullable :: (Ord \ nt) \Rightarrow Grammar \ nt \ t \rightarrow Set \ nt
nullable = nullable' S.empty
nullable' :: (Ord \ nt) \Rightarrow Set \ nt \rightarrow Grammar \ nt \ t \rightarrow Set \ nt
nullable' set grammar = set'' where
  set'' = \mathbf{if} \ nulls \equiv set \ \mathbf{then} \ set \ \mathbf{else} \ set'
  set' = nullable' nulls (S.fold drop grammar nulls)
  nulls = S.fromList \circ map\ nonterminal \circ filter\ isNull\ \$\ grammar
isNull :: Production \ nt \ t \rightarrow Bool
isNull (Production \_Empty) = True
isNull \ \_ = False
simpleGrammar :: Grammar String String
simpleGrammar = [a] where
  a = Production "A" (Term "ab" Empty)
simpleGrammar2 :: Grammar String String
simpleGrammar2 = [a, a', b, b', c] where
  a = Production \ "A" \ (\textit{Term "ab" } \textit{Empty})
```

```
a' = Production "A" \ Empty b = Production "B" \ (NonT "A" \ (NonT "A" \ Empty)) b' = Production "B" \ (NonT "A" \ (Term "b" \ Empty)) c = Production "C" \ (Term "cdef" \ Empty)
```

5.2 First

In this section, we compute the first set for a context-free grammar.

```
module First (first) where
import ContextFreeGrammar
import Control.Monad
import Control.Monad.State
import Data.Functor
import Data.List
import qualified Data.Map as M
import Data.Maybe
import Nullable
import ScanAndParse
import qualified Data.Set as S
import Test.HUnit hiding (State)
```

first is the interface function exported for general use. Given a Grammar, first computes the First Set for each Production, and returns the Sets in a Map from a Non-terminal to it's First Set.

```
 first :: (Ord \ nt, Ord \ t) \\ \Rightarrow Grammar \ nt \ t \\ \rightarrow M.Map \ nt \ (S.Set \ (Terminal \ t)) \\ first \ g = firsts \circ execState \ state \circ FS \ M.empty \circ nullable \ g \ \mathbf{where} \\ state = mapM \ first' \circ concat \circ replicate \ (length \ g) \ g
```

The FirstState Data Type stores the map of Sets of First Terminals that is modified and returned at the end of a call to first. It also stores the set of Non-terminals which are nullable.

```
data FirstState \ nt \ t = FS \ \{

firsts :: M.Map \ nt \ (S.Set \ (Terminal \ t)),

nulls :: S.Set \ nt

\}
```

```
type Environment nt \ t \ a = State \ (FirstState \ nt \ t) \ a
```

first' does the work for the first function. Given a production, first' will calculate the set of first terminals and store it in the implicit FirstState.

```
first':: (Ord nt, Ord t) \Rightarrow Production nt t \rightarrow Environment nt t ()
first' (Production nt rhs) = \operatorname{\mathbf{do}}
fs \leftarrow get
let mp = firsts fs
case rhs of

Empty \rightarrow case M.lookup nt mp of

Nothing \rightarrow put fs {firsts = M.insert nt (S.singleton Epsilon) mp}

Just _{-} \rightarrow put fs {firsts = M.adjust (S.insert Epsilon) nt mp}

_{-} \rightarrow \operatorname{\mathbf{do}}
sets \leftarrow firstRHS rhs
let s = fromMaybe S.empty (M.lookup nt mp)
put fs {firsts = M.insert nt (S.unions (s:sets)) mp}
```

firstRHS is a helper function which, given a RHS will return the first sets of every terminal and non-terminal until a non nullable terminal/nonterminal is found.

```
\begin{array}{l} \mathit{firstRHS} :: (\mathit{Ord}\ \mathit{nt}, \mathit{Ord}\ \mathit{t}) \\ \Rightarrow \mathit{RHS}\ \mathit{nt}\ \mathit{t} \\ \rightarrow \mathit{Environment}\ \mathit{nt}\ \mathit{t}\ [\mathit{S.Set}\ (\mathit{Terminal}\ \mathit{t})] \\ \mathit{firstRHS}\ \mathit{Empty} = \mathit{return}\ [\mathit{I} \\ \mathit{firstRHS}\ (\mathit{Term}\ \mathit{y}\ \_) = \mathit{return}\ [\mathit{S.singleton}\circ \mathit{Terminal}\ \$\ \mathit{y}] \\ \mathit{firstRHS}\ (\mathit{NonT}\ \mathit{y}\ \mathit{ys}) = \mathbf{do} \\ \mathit{nlls} \leftarrow \mathit{gets}\ \mathit{nulls} \\ \mathbf{case}\ \mathit{S.member}\ \mathit{y}\ \mathit{nlls}\ \mathbf{of} \\ \mathit{True} \rightarrow \mathbf{do} \\ \mathit{set} \leftarrow \mathit{getFirsts}\ \mathit{y} \\ \mathit{sets} \leftarrow \mathit{firstRHS}\ \mathit{ys} \\ \mathit{return}\ (\mathit{set}:\mathit{sets}) \\ \mathit{False} \rightarrow (:[]) < \$ > \mathit{getFirsts}\ \mathit{y} \end{array}
```

getFirsts is a helper function which given a Non-terminal will return its current first set.

```
getFirsts :: Ord \ nt \Rightarrow nt \rightarrow Environment \ nt \ t \ (S.Set \ (Terminal \ t))
getFirsts \ nt = \mathbf{do}
```

```
set \leftarrow (M.lookup\ nt) < \$ > gets\ firsts
  case set of
     Nothing \rightarrow return \ S.empty
     \textit{Just set} \rightarrow \textit{return set}
 {- BEGIN TESTS - -}
makeTestM :: (Eq \ a, Show \ a)
   \Rightarrow String
   \rightarrow FilePath
   \rightarrow String
   \rightarrow a
   \rightarrow (Grammar String String \rightarrow a)
   \rightarrow Test
makeTestM name file forF e f = TestLabel name \circ TestCase \$ do
  grammar \leftarrow fmap \ sparse \circ readFile \ file
  assertEqual\ for F\ e\ (f\ grammar)
testFirst = makeTestM "testFirst"
  "tests\\test1.txt"
  "for first with test1"
  expected
  first where
  expected = M.fromList [("A", S.singleton \circ Terminal \$ "a"),
    ("B", S.fromList [Terminal "b",
        Terminal "a",
       Epsilon]),
     ("C", S.fromList [Terminal "a",
        Terminal "b",
       Epsilon]),
       ("D", S.fromList [Terminal "a",
          Terminal "b",
          Epsilon])]
testFirst2 = makeTestM "testFirst2"
  "tests\\39.txt"
  "for first with 39"
  expected
  first where
  expected = M.fromList [("T", S.fromList [Epsilon, Terminal "a", Terminal "b"]),
     ("R", S.fromList [Epsilon, Terminal "b"])]
tests = TestList [testFirst,
  testFirst2\, \rceil
```

```
runTests :: IO\ Counts
runTests = runTestTT\ tests
do\ TestsPass :: IO\ Bool
do\ TestsPass = \mathbf{do}
counts \leftarrow runTests
\mathbf{let}\ errs = errors\ counts
fails = failures\ counts
return\ \$\ (errs \equiv 0) \land (fails \equiv 0)
\{-\ END\ TESTS - -\}
```

5.3 Follow

In this section, we implement a function *follow* which calculates the follow set for our data structure of production grammars.

```
module Follow (follow) where
import ContextFreeGrammar
import Control.Monad.State
import Data.Functor
import qualified Data.Map as M
import Data.Maybe
import qualified Data.Set as S
import First
import Nullable
import ScanAndParse
import Test.HUnit hiding (State)
```

A GrammarState holds data that the follow' function requires to work.

```
data GrammarState \ nt \ t = GS \ \{
grammar :: Grammar \ nt \ t,
follows :: M.Map \ nt \ (S.Set \ (Terminal \ t)),
firsts :: M.Map \ nt \ (S.Set \ (Terminal \ t))
\}
type Environment \ nt \ t \ a = State \ (GrammarState \ nt \ t) \ a
```

follow is the interface function exported for general use. Given a Grammar, follow computes the Follow Set for each Production, and returns the Sets in a Map from a Non-terminal to it's Follow Set.

```
 follow :: (Ord \ nt, Ord \ t) \\ \Rightarrow Grammar \ nt \ t \\ \rightarrow M.Map \ nt \ (S.Set \ (Terminal \ t)) \\ follow \ [] = M.empty \\ follow \ g@((Production \ s \ rhs) : ps) = fs \ \mathbf{where} \\ state = mapM \ follow' \circ concat \circ replicate \ (length \ g) \ \$ \ g \\ fs = follows \circ execState \ state \circ GS \ g \ initial \circ first \ \$ \ g \\ initial = M.singleton \ s \ (S.singleton \ EOF)
```

follow' is where the main work of the follow function is done. For a given production, follow' will add an entry into the GrammarState passed along. This function is meant to be mapM'd across the Grammar you want to compute the follow sets of.

```
follow' :: (Ord \ nt, Ord \ t)
   \Rightarrow Production \ nt \ t
   \rightarrow Environment\ nt\ t\ ()
follow' (Production \ a \ \_) = \mathbf{do}
  fllostate \leftarrow qet
  let g = grammar fllostate
        fllow = follows fllostate
         ps = filter (elem \ a \circ nonTerminals \circ rhs) \ g
   sets \leftarrow forM \ ps \ \lambda(Production \ x \ rhs) \rightarrow \mathbf{do}
      case after a rhs of
         Empty \rightarrow return \circ from Maybe \ S.empty \circ M.lookup \ x \ fillow
         Term\ t \longrightarrow return \circ S.singleton \circ Terminal \ t
         NonT \ nt \ rest \rightarrow \mathbf{do}
            let frstb = (firsts \ filostate)M.! nt
              fllowx = \mathbf{case} \ S.member \ Epsilon \ frstb \ \mathbf{of}
                  False \rightarrow S.empty
                  True \rightarrow from Maybe\ S.empty \circ M.lookup\ x\ \$ fllow
            return \$ S.union (S.delete Epsilon frstb) (fllowx)
  let s = fromMaybe\ S.empty\ (M.lookup\ a\ fllow)
         newS = S.union \ s \ (S.unions \ sets)
  put\ fllostate\ \{follows = M.insert\ a\ newS\ fllow\}
```

after is a helper function which removes all Terminals and Non-terminals from a RHS until a specific Non-terminal is reached. Then the rest of the RHS is returned.

```
after :: (Eq \ nt) \Rightarrow nt \rightarrow RHS \ nt \ t \rightarrow RHS \ nt \ t
 after \ nt \ Empty = Empty
```

```
after nt (Term t rhs) = after nt rhs
after nt (NonT nt2 rhs) = if nt \equiv nt2
  then rhs else after nt rhs
 {- BEGIN TESTS - -}
makeTestM :: (Eq \ a, Show \ a)
   \Rightarrow String
   \rightarrow FilePath
   \rightarrow String
   \rightarrow a
   \rightarrow (Grammar String String \rightarrow a)
   \rightarrow Test
makeTestM name file forF e f = TestLabel name \circ TestCase \$ do
  grammar \leftarrow sparse < \$ > readFile file
  assertEqual\ forF\ e\ (f\ grammar)
testFollow = makeTestM "testFollow"
  "tests\\39.txt"
  "for first with 39"
  expected
  follow where
  expected = M.fromList [("R", S.fromList [EOF, Terminal "c"]),
     ("T", S.fromList [EOF, Terminal "c"])]
tests = TestList [testFollow]
runTests :: IO\ Counts
runTests = runTestTT tests
doTestsPass :: IO Bool
doTestsPass = \mathbf{do}
  counts \leftarrow runTests
  let \ errs = errors \ counts
     fails = failures \ counts
  return \$ (errs \equiv 0) \land (fails \equiv 0)
 {- END TESTS - -}
```

6 Generating a Parse Table

In this section we generate a parse table for a given grammar, assuming it has been properly scanned, parsed, and thoroughly cleansed.

module Table where

```
\mathbf{import}\ \mathit{ContextFree}\ \mathit{Grammar}
import Filterable
import Nullable
import First
import Follow
import System. Environment
import Data.List as L
import qualified Data. Map as M
import qualified Data. Set as S
import ScanAndParse
import BadHygiene
type Table = M.Map (String, String) (Production String String)
foo1 :: Grammar String String
foo1 = [a, b, c, d] where
  a = Production "A" (Term "a" Empty)
  b = Production "B" (NonT "B" Empty)
  c = Production "C" (Term "a" (NonT "B" Empty))
  d = Production "D" (NonT "B" (Term "a" Empty))
getFeature feature productions = loop productions []
  where
    loop [] acc = sort \circ nub \circ concat \$ acc
    loop\ ((Production\ s\ rhs): xs)\ acc = loop\ xs\ ((feature\ rhs): acc)
  -- need to generate S' \to S$
getNewStart :: Grammar String String 
ightarrow Grammar String String
getNewStart[] = error "getNewStart run on empty grammar --- a new low point"
getNewStart \ g@(Production \ nt \ rhs: ps) =
    (Production\ (nt + ",")\ (NonT\ nt\ (Term\ "\$"\ Empty))): g
isRHSNullable :: Ord \ a \Rightarrow S.Set \ a \rightarrow RHS \ a \ t \rightarrow Bool
isRHSNullable \_Empty = True
isRHSNullable \_ (Term \ t \_) = False
isRHSNullable \ m \ (NonT \ nt \ rhs) =
  if S.member nt m then
    isRHSNullable m rhs
  else False
firstRHS = Empty = S.empty
firstRHS \ i \ (Term \ t \ rhs) = S.singleton \ t
firstRHS \ i \ (NonT \ nt \ rhs) =
```

```
S.union firstsnt (firstRHS i rhs)
         else
           firstsnt
         where firstsnt = ((firsts\ i)M.!\ nt)
      data GrammarInfo = GI {
        firsts :: M.Map String (S.Set String),
         follows :: M.Map String (S.Set String),
         nulls :: S.Set String
      } deriving Show
    A production N \to \alpha is in the table (N,a) iff a is in first (\alpha) \lor (\text{nullable}(\alpha))
\wedge a is in follow(\alpha)). This condition is readily translateable into our code:
      validEntry\ qi\ (Production\ nterm\ \alpha)\ term=firstalpha\ \lor\ (nullablea\ \land\ follown)
         where
           firstalpha = S.member (term) (firstRHS qi \alpha)
           nullablea = isRHSNullable (nulls qi) \alpha
           follown = S.member (term) ((follows qi)M.!nterm)
    We'll also need some conversion from the Terminal data types used by
the first and follow functions:
      to Terminal [] = []
      to Terminal ("\$" : ss) = EOF : to Terminal ss
      toTerminal(s:ss) = Terminal(s:toTerminal(ss))
      from Terminal' :: Terminal String \rightarrow String
      from Terminal' Epsilon = ""
      from Terminal' EOF
      from Terminal' (Terminal t) = t
      from Terminal Set \ ts = S.map \ from Terminal' \ ts
    Lastly, we build a table for an LL(1) parser, returning nothing if there
is more than one production per entry in the table.
      buildTable' :: GrammarInfo \rightarrow Grammar\ String\ String \rightarrow [String] \rightarrow Table \rightarrow Maybe\ Table
      buildTable' \_[] \_acc = Just\ acc
      buildTable' gi (p@(Production\ nt\ \_):ps) terms\ acc = \mathbf{case}\ fromList\ acc\ kvs\ \mathbf{of}
         Nothing \rightarrow Nothing
         Just a \rightarrow buildTable' qi ps terms a
         where
```

if (S.member nt (nulls i)) then

```
valids = \\ map \ fst \circ L. \ filter \ snd \circ map \ (\lambda t \to (t, validEntry \ gi \ p \ t)) \ \$ \ terms \\ kvs = map \ (\lambda v \to ((nt, v), p)) \ valids \\ from List :: Table \to \\ [((String, String), Production \ String \ String)] \to Maybe \ Table \\ from List \ acc \ [] = Just \ acc \\ from List \ acc \ ((k@(nt, t), p) : ks) = \mathbf{case} \ M. lookup \ k \ acc \ \mathbf{of} \\ Nothing \to from List \ (M. insert \ k \ p \ acc) \ ks \\ Just \ \_ \to Nothing \\ build Table \ grammar = \\ \mathbf{let} \ terms = get Feature \ terminals \ grammar \ \mathbf{in} \\ \mathbf{let} \ gi = GI \ (M. map \ from TerminalSet \ \$ \ first \ grammar) \ (M. map \ from TerminalSet \ \$ \ folionet \ table = build Table' \ gi \ grammar \ terms \ M. empty \ \mathbf{in} \\ table
```

6.1 Table-driven Parser module

In this section we provide a simple table-driven parser function, *parseWithTable*. It takes a start non-terminal symbol, a table created in the *Table* module, an input string, and returns a boolean indicating whether the string was successfully parsed or not.

Since we are just creating a parser, but do not know the intended use of the parser, we decided to simply return a boolean. If necessary, it could be easily modified for other accommodations. n Lastly, we decided to emit Haskell code as a module, to run on a given text file as input for that the context free grammar it was generated for.

```
module Parser where
import ContextFreeGrammar
import Data. Char
import System. Environment
import qualified Data. Map as M
import ScanAndParse
import BadHygiene
import Table
getStart :: Grammar String String \rightarrow String
getStart ((Production nt _): _) = nt
rhsString =
```

```
"rhsToList :: RHS String String -> [String]\n" ++
  "rhsToList Empty = []\n" ++
  "rhsToList (Term t rhs) = t : rhsToList rhs\n" ++
  "rhsToList (NonT nt rhs) = nt : rhsToList rhs"
parseWithTableString =
  "\nparseWithTable start t s = parse' [start] s where\n" +
    parse' [] cs = True\n" ++
    parse' (top:rest) s@(c:cs)\n" +
       | isUpper (head top) = case M.lookup (top,[c]) t of\n" ++
         Nothing -> False\n" ++
         Just (Production _ rhs) -> parse' (rhsToList rhs ++ rest) s\n" ++
       | c == head top = parse' rest cs\n" #
       | otherwise = parse' rest s\n"
generateStrong\ grammar\ table = \mathbf{do}
 putStrLn "module LLParser (parse) where "
 putStrLn "import qualified Data.Map as M"
 putStrLn "import ContextFreeGrammar"
 putStrLn "import Data.Char"
 putStr "\ntable = M."
 putStrLn \circ show \$ table
 putStrLn parseWithTableString
 putStrLn rhsString
 putStr \$ "parse s = parseWithTable " ++ "\"" ++ (getStart\ grammar) ++ "\"" ++ " tab
```

7 Main module

The main module puts everything together, takes a textual representation of a context-free grammar as input, scans, parses, and performs the rest of the duties that are required.

```
module Main where
import ContextFreeGrammar
import ScanAndParse
import BadHygiene
import Table
import Parser
import System.Environment
main = do
```

```
[file] \leftarrow getArgs \\ contents \leftarrow readFile \ file \\ \textbf{let} \ grammar = sparse \ contents \\ \textbf{let} \ grammar' = eliminate Useless \circ getNewStart \$ \ grammar \\ \textbf{let} \ table = buildTable \ grammar' \\ \textbf{case} \ table \ \textbf{of} \\ Nothing \rightarrow error \ "Not \ done" \\ Just \ t \rightarrow generateStrong \ grammar' \ t
```

8 Conclusion

This concludes our implementation. We provide some sample output of a generated binary run on a context free grammar specified in a text file, 39.txt.