ocaml 第四期レポート課題

高口 奨一郎 学籍番号:6319045

2020年7月30日

1 解いた問題

問題1-8まで

2 問題1

字句解析部の実装を次に示す. 省略してある部分にコードを付加し, ''module Lexer = struct ··· end''で括ってプログラムを完成させなさい. さらに, 四則演算同様, gettoken を繰り返し呼び出す ''run'' 関数を実装し, 振舞を確かめなさい.

2.1 プログラムソース

```
1 \mod \text{ule Lexer} = \text{struct}
2
3 \text{ module } P = Printf
4
5 type token = CID of string | VID of string | NUM of string
                 | TO | IS | QUIT | OPEN | EOF | ONE of char
6
7
  exception End_of_system
9
  let JSTREAM = ref stdin
10
11
   let ch = ref
12
13
   let read () = match !ch with [] -> input_char !JSTREAM
14
                           | h :: rest \rightarrow (ch := rest; h)
15
16
  let unread c = ch := c :: ! ch
17
18
```

```
19 let lookahead () = try let c = read () in unread c; c
20
                                               with End_of_file -> '$
21
  let rec integer i =
   (* 文字列として数字を構成 *)
23
    let c = lookahead () in
       if (c >= '0' \&\& c <= '9') then
24
              integer (i^(Char.escaped (read ())))
25
       else i
26
27
   and identifier id =
29
    let c = lookahead () in
       if ((c >= 'a' \&\& c <= 'z') || (c >= 'A' \&\& c <= 'Z') ||
30
          (c >= '0' \&\& c <= '9') \mid |c == '_-') \text{ then}
31
              identifier (id^(Char.escaped (read ())))
32
33
       else id
34
   and native\_token () =
35
36
    let c = lookahead () in
      if (* CID に対する識別子および予約語 *)
37
         (c >= 'a' \&\& c <= 'z') then
38
             let id = identifier "" in
39
                 match id with
40
                  " i s " -> IS
41
                 |" quit" -> QUIT
42
                 |" open" → OPEN
43
                 | ' ' e o f ' ' ' -> EOF
44
                 | \_ \rightarrow CID (id)
45
      else if (* VID に対する識別子 *)
46
         (c \ge 'A' \&\& c \le 'Z') then VID (identifier "")
47
```

```
else if (c \ge 0') && c \le 9') then NUM (integer "")
48
      else if (* :- を認識して TO を返す *)
49
          (c=':') then (let sub1 = read () in let sub2 = read ()
50
                          in if sub2 = '-' then TO
51
52
                              else ONE (sub1))
       else ONE (read ())
53
54
55
   and gettoken () =
56
   try
   let token = native_token () in
57
58
     match token with
        ONE ' ' \rightarrow gettoken ()
59
      | ONE ' \setminus t' -> gettoken ()
60
        ONE ' n' \rightarrow gettoken ()
61
      _ -> token
62
   with End_of_file -> EOF
63
64
65
   let print_token tk =
  match tk with
66
   (CID i) \rightarrow P.printf "CID(\%s)" i
67
     (VID i) -> P. printf "VID(%s)" i
68
     (NUM i) \rightarrow P.printf "NUM(\%s)" i
69
     (TO) -> P. printf ":-"
70
     (QUIT) -> P. printf "quit"
71
     (OPEN) -> P. printf "open"
72
     (IS) -> P. printf "is"
73
     (EOF) -> P. printf "eof"
74
     (ONE \ c) \rightarrow P.printf "ONE(\%c)" \ c
75
76
```

2.2 省略部分のプログラムの説明

2.2.1 21 **行目の関数** integer

この関数は入力された数字を文字列として返す関数である。まず lookahead を用いて文字を見る。その文字が数字(0 から 9)であれば read を用いてその値を Char.escaped に与えることで文字列に変更する。また関数 integer を再帰的に用いることでバラバラな文字として認識されていた数字をひとつにまとめる。なお、入力された文字が数字でない場合にはそのままi を返す。

2.2.2 37 行目の関数 native __ token の内部

この関数は関数 identifier を用いてそれぞれの文字のトークンが何にあたるかを判別していく。 まず、先頭が小文字で始まるものについて考える。小文字で始まる文字列のとき、is, quit, open., eof, CID のいずれかであるのでmatch 文を用いてそれぞれにあったトークン IS,QUIT,OPEN,EOF,CID(id)を返す。次に先頭が大文字のアルファベットで始まるものについては VIDトークン、先頭が数字で始まるものについては NUMトークンを返す。先頭が:であるときには:を読み飛ばしたあとで:の次の文字を read で見てーであれば TO を返し、そうでなければただのコロンとして ONE(sub1) を返す。

2.3 実行結果

以下のように正しく動作していることがわかる。

```
val run: unit -> unit = <fun>
# open Lexer;;
# gettoken();;
:-
-: Lexer.token = TO
# settoken();;
-: Lexer.token = ONE '-'
# settoken();;
-: Lexer.token = ONE '-'
# settoken();;
open
-: Lexer.token = OPEN
# settoken();;
quit
-: Lexer.token = BOF
# settoken();;
eof
-: Lexer.token = EOF
# settoken();;
is
-: Lexer.token = EOF
# settoken();;
suj
-: Lexer.token = CID "suj"
# settoken();;
SUJ
-: Lexer.token = VID "SUJ"
# settoken();;
**
-: Lexer.token = ONE '+'
# settoken();;
**
-: Lexer.token = NUM "8"
# settoken();;
**
-: Lexer.token = NUM "8"
# settoken();;
-: Lexer.token = TO
#
```

```
# run();;
1+2;
NUM(1)
ONE(+)
NUM(2)
ONE(;)
father(masuo,tara).
CID(father)
ONE(()
CID(masuo)
ONE(,)
CID(tara)
ONE())
ONE())
ONE())
ONE())
ONE())
ONE()
```

文法中の、terms、args は、左再帰の形をしているので、右再帰に変換する必要がある。修正した文法を示しなさい。左再帰を取り除いた文法から作成した構文解析器は、次のようになる。

3.1 左再帰を除いた文法

新しく terms'、args' を定義することで左再帰を除き、右再帰に変換することができる。

3.1.1 terms

```
terms \rightarrow term terms'
terms' \rightarrow "," term terms'
terms' \rightarrow
```

3.1.2 args

```
args \rightarrow expr args'
args' \rightarrow "," expr args'
args' \rightarrow
```

構文解析器プログラムにある空欄を埋めて, プログラムを完成させなさい.

4.1 プログラムソース

```
1 module Parser = struct
2
3 \text{ module } L = Lexer
4
  (* module E = Evaluator *)
6
  let tok = ref (L.ONE', ')
8
  let getToken () = L.gettoken ()
10
   let advance () = (tok := getToken(); L.print_token (!tok))
11
12
13
   exception Syntax_error
14
15
   let error () = raise Syntax_error
16
   let check t = match ! tok with
17
       L.CID \longrightarrow if (t = (L.CID "")) then () else error()
18
      | L.VID | -> if (t = (L.VID "")) then () else error()
19
      | L.NUM | \rightarrow if (t = (L.NUM"")) then () else error()
20
      | tk -> if (tk=t) then () else error()
21
22
23 let eat t = (check t; advance())
24
```

```
25 let rec clauses () = match !tok with
       L.EOF \rightarrow ()
26
      | _ -> (clause(); clauses())
27
28
29
   and clause () = match ! tok with
       L.ONE '(' -> (term(); eat(L.ONE '.'))
30
      _ -> (predicate(); to_opt(); eat(L.ONE '.'))
31
32
   and to_opt() = match !tok with
       L.TO \rightarrow (eat(L.TO) ; terms())
34
    | _ -> ()
35
36
   and command() = match !tok with
37
       L.QUIT \rightarrow exit 0
38
      | L.OPEN \rightarrow (eat(L.OPEN);
39
         match ! tok with
40
             L.CID s \rightarrow (eat(L.CID ""); check (L.ONE '.');
41
             L.ISTREAM := open_in (s^*.pl^*);
42
             advance(); clauses(); close_in (!L._ISTREAM))
43
           | _ -> error())
44
      _ -> (term(); check(L.ONE '.'))
45
46
   and term() = match !tok with
47
       L.ONE '(' -> (eat(L.ONE '(') ; term() ; eat(L.ONE ')'))
48
      | _ -> predicate()
49
50
   and terms() = (term() ; terms'())
51
52
53 and terms' () = match !tok with
```

```
L.ONE ',' -> (eat(L.ONE ',') ; term() ; terms'())
54
      | _ -> ()
55
56
   and predicate() = (eat(L.CID ""); eat(L.ONE '(');
57
                                             args(); eat(L.ONE ')'))
58
59
   and args() = (expr(); args'())
61
   and args'() = match !tok with
       L.ONE ', ' -> (eat(L.ONE ', ') ; expr() ; args '())
63
     | _ -> ()
64
65
   and expr() = match ! tok with
       L.ONE'(' \rightarrow (eat(L.ONE'(') ; expr() ; eat(L.ONE')'))
67
      | L.ONE '[' -> (eat(L.ONE '[') ; list() ; eat(L.ONE ']'))
68
      | L.CID s \rightarrow (eat(L.CID "") ; tail\_opt())
69
      | L.VID s \rightarrow (eat(L.VID ""))
70
      | L.NUM n \rightarrow (eat(L.NUM ""))
71
      | _ -> error()
72
73
   and tail_opt() = match !tok with
74
       L.ONE '(' -> (eat(L.ONE '(') ; args() ; eat(L.ONE ')'))
75
     | _ -> ()
76
77
78
   and list() = match !tok with
       L.ONE '] '\rightarrow ()
79
      | - > (expr(); list_opt())
80
81
82 and list_opt() = match ! tok with
```

```
L.ONE '| ' -> (eat(L.ONE '| '); id())
83
       | L.ONE ', ' -> (eat(L.ONE ', '); list())
84
       | _ -> ()
85
86
87
   and id() = match ! tok with
       L.CID _ -> (eat(L.CID ""))
88
       |L.VID _ -> (eat(L.VID ""))
89
      |L.NUM _ -> (eat(L.NUM ""))
90
      _ -> error()
91
92
93 end
94
95 let rec run() =
96 print_string "?- ";
97 while true do
98 flush stdout; Lexer. JSTREAM := stdin;
99 Parser.advance(); Parser.command(); print_string "\n?-"
100 done
```

4.2 省略部分のプログラムの説明

4.2.1 33 **行目の関数** to __ opt

match 文を用いて入力が TO であれば eat した上で terms を実行、そうで なければ何もしない。

4.2.2 47 行目の関数 term

match 文を用いて入力の先頭の文字が'('であれば'('を eat した上で term の実行、')'の eat を行う。そうでなければ predicate を実行する。

4.2.3 51 行目の関数 terms

左再帰を除くため新たに terms'を定義する。問2で考えた文法の規則に従ってプログラムする。terms では term を実行した後に terms'を、terms'では match 文を用いて入力の先頭の文字が'('であれば'('を eat した上で term の実行、terms'の実行を行う。そうでなければ何もしない。

4.2.4 57 **行目の関数** predicate

CID、'('の eat、args の実行、')'の eat を順に行う。

4.2.5 60 **行目の関数** args

terms と同じようにして問 2 の文法を参考にしてプログラムを設定する。 args では expr を実行した後に args'を、args'では match 文を用いて入力の 先頭の文字が'('であれば'('を eat した上で expr の実行、args'の実行を行う。そうでなければ何もしない。

4.2.6 66 行目の関数 expr

match 文を用いて入力の先頭の文字が'('であれば'('を eat した上で expr の実行、')'の eat を行う。入力の先頭の文字が'['であれば'['を eat した上で list の実行、']'の eat を行う。入力が CID であれば CID を eat した上で tail __ opt を実行、入力が VID、NUM であれば VID の eat、NUM の eat をそれぞれ実行する。そうでなければエラーメッセージを返す。

4.2.7 74 **行目の関数** tail __ opt

match 文を用いて入力の先頭の文字が'('であれば'('を eat した上で args の実行、')'の eat を行う。そうでなければ何もしない。

4.2.8 87 行目の関数 id

match 文を用いて入力が CID であれば CID の eat、入力が VID であれば VID の eat、入力が NUM であれば NUM の eat を行う。そうでなければエラーメッセージを返す。

isono.pl を用いて振舞いを確認しなさい. このように正常に動作している

ことがわかる。

上記で示した文法とプログラムは、質問として 1 つの述語しか与えることができない.「、」で区切って複数の述語を記述できるようにしなさい.

6.1 アルゴリズムの説明

述語を与える predicate を","で区切ったときにもう一度再帰的に呼び出せばよい。

6.2 プログラム

6.3 プログラムの説明

関数 predicate の最後に関数 predicate'を加える。関数 predicate に加えて新しい関数 predecate'を定義する。関数 predicate'は match 文を用いて入力の先頭の文字が','であれば','を eat した上で関数 predicate と同じ操作を行う。そうでなければ入力ががこれ以上ないということなので何もしない。

6.4 実行結果

以下のように正しく動作していることがわかる。

```
?- grandfather(namihei,tara),father(namihei,katsuo).
CID(grandfather)ONE(()CID(namihei)ONE(,)CID(tara)ONE())ONE(,)CID(father)ONE(()CI
ONE(,)CID(katsuo)ONE())ONE(.)
```

文法中にエラーが生じた際に、エラーメッセージを印字するようにしなさい. エラーメッセージには、エラーが生じた箇所の行番号を含むこと. 元のプログラムでは、文法エラーに対して例外を発生させるので、この例外を捕まえて処理する必要があるだろう. OCaml で例外処理を記述するには、次のようにする.

7.1 アルゴリズムの説明

改行を示すトークンを判別するたびにカウントを増やし、文法エラーが生じたタイミングで箱に格納しておいたカウントをエラーが発生した行数としてエラーメッセージと共に出力すればよい。

7.2 プログラムの説明

```
1 let count = ref 1
2
3 let sub () = count := !count + 1
4
5 let reset() = count := 0
```

Lexer モジュール内に以上の3つの関数を定義する。関数 count は行番号をカウントする関数である。関数 sub は呼び出されるたびに count の値を 1 増加させる。関数 reset は count の初期値を0 にする。

1 and gettoken () =
2 try
3 let token = native_token () in
4 match token with
5 ONE ' '-> gettoken ()
6 | ONE '\t' -> gettoken ()
7 | ONE '\n' -> sub() ; gettoken ()

```
9 with End_of_file -> EOF
  改行を判別するたびに count の値を 1 増加させる関数 sub を呼び出して
  いる。
 let rec run() =
     print_string "?- ";
2
3
     while true do
       flush stdout; Lexer. JSTREAM := stdin;
4
       Parser.advance(); try Parser.command() with Parser.Syntax
5
       print_int(!Lexer.count); print_string 行目"Syntax_error";
6
       print_string "\n?-"
```

Parser.command() によって発生したエラーを捕まえて print によって?行 目 Syntax __ error と出力されるように設定した。

7.3 実行結果

| _ -> token

Lexer.reset ();

8

7

8

このように文法エラーが生じた行とエラーメッセージを返す。

```
father(X,Y),
nother.
CID(father)ONE(()VID(X)ONE(,)VID(Y)ONE())ONE(,)CID(mother)ONE(.)2行目Syntax_erro
  father(X,Y),
ID(father)ONE(()VID(X)ONE(,)VID(Y)ONE())ONE(,)CID(mo)CID(th)3行目Syntax_error
  (father)ONE(()VID(X)ONE(,)VID(Y)ONE())ONE(,)CID(mother)ONE(()VID(X)ONE(.)3行
  father(X,Y).
 .
D(father)ONE(()VID(X)ONE(,)VID(Y)ONE())ONE(,)CID(mother)ONE(()VID(X)ONE(.)3行
```

SProlog の文法に算術式を加え、それに基づいて構文解析器を拡張しなさい。付加する生成規則は次のようになる。元の文法において、expr の生成規則の右辺を、arithmexp で置き換えるとよい。

8.1 アルゴリズムの説明

生成規則に従ってプログラムを構成すればよい。ただし、生成規則をそのまま反映させると arithmexp と arithmterm は左再帰になっているので右再帰に変換した文法を考える。以下に右再帰に変換した文法を示す。 arithmfactor に関しては左再帰を含んでいないので生成規則通りにプログラム記述する。

```
8.1.1 arithmexp
arithmexp → arithmterm arithmexp'
arithmexp' → "+" arithmterm arithmexp'
arithmexp' → "-" arithmterm arithmexp'
arithmexp" →
8.1.2 arithmterm
arithmterm → arithmfactor arithmterm'
arithmterm' → "*" arithmfactor arithmterm'
arithmterm' → "/' arithmfactor arithmterm'
```

8.2 プログラム

arithmterm' \rightarrow

```
1 and arithmexp() = (arithmterm(); arithmexp'())
2
3 and arithmexp'() = match !tok with
4 L.ONE '+' -> (eat(L.ONE '+'); arithmterm(); arithmexp'())
5 |L.ONE '-' -> (eat(L.ONE '-'); arithmterm(); arithmexp'())
```

```
6 | _ -> ()
7
8 and arithmterm() = (arithmfactor(); arithmterm'())
9
   and arithmterm' () = match !tok with
    L.ONE '*' -> (eat(L.ONE '*') ; arithmfactor() ; arithmterm'())
11
   |L.ONE '/' ->(eat(L.ONE '/'); arithmfactor(); arithmterm'())
12
   | _ -> ()
13
14
15
   and arithmfactor() = match !tok with
    L.ONE '(' \rightarrow (eat(L.ONE '(') ; arithmexp() ; eat(L.ONE ')'))
16
    L.ONE '-' -> (eat(L.ONE '-') ; arithmexp())
17
    L.ONE '[' -> (eat(L.ONE '[') ; list() ; eat(L.ONE ']'))
18
   | L.CID s \rightarrow (eat(L.CID "") ; tail_opt())
19
   | L.VID s \rightarrow (eat(L.VID ""))
20
     L.NUM n \rightarrow (eat(L.NUM ""))
21
22
     _ -> error()
```

8.3 プログラムの説明

関数 arirhmexp は arithmterm() を実行した後に arithmexp'() を実行する。関数 arithmexp'は match 文を用いて先頭の文字が+であれば'+'を eat した上で arithmterm() を実行し、arithmexp'() を実行する。先頭の文字が一であれば'ー'を eat した上で arithmterm() を実行し、arithmexp'() を実行する。そうでなければ何もしない。ここがループが停止するところである。

関数 arirhterm は arithmfactor() を実行した後に arithmterm'() を実行する。関数 arithmterm'は match 文を用いて先頭の文字が*であれば'*'を eat した上で arithmfactor() を実行し、arithmterm'() を実行する。先頭の文字が/であれば'/'を eat した上で arithmfactor() を実行し、arithmterm'()

を実行する。そうでなければ何もしない。ここがループが停止するところで ある。

関数 arithmfactor は match 文を用いて先頭の文字が'('であれば'('を eat し、arithmexp()を実行、')'を eat する。先頭の文字が' – 'であれば' – 'を eat し、arithmexp()を実行、先頭の文字が' ['であれば' ['を eat し、list()を実行、']'を eat する。先頭の文字が CID であれば CID を eat し、tail __opt()を実行、先頭の文字が VID、NUM であれば VID、NUM を eat する。そうでなければエラーを返す。

8.4 実行結果

```
1 and expr() (*= match !tok with
2     L.ONE '(' -> (eat(L.ONE '(') ; expr() ; eat(L.ONE ')'))
3     | L.ONE '[' -> (eat(L.ONE '[') ; list() ; eat(L.ONE ']'))
4     | L.CID s -> (eat(L.CID "") ; tail_opt())
5     | L.VID s -> (eat(L.VID ""))
6     | L.NUM n -> (eat(L.NUM ""))
7     | - -> error()
8
9 and arithmexp() *)=
```

この部分をコメントアウトする。 以下のように正しく動作していることがわかる。

```
val id: unit -> unit
end
val run: unit -> unit = <fun>
# open Parser;;
# run();;
?- mother (x+y).
CID(mother)ONE(()CID(x)ONE(+)CID(y)ONE())ONE(.)
?- father ([3]).
CID(father)ONE(()ONE([)NUM(3)ONE(])ONE())ONE(.)
?- ■
```

変数への代入を記述できるようにしなさい.必要な生成規則は次のようになる.

9.1 アルゴリズムの説明

生成規則に従って term に新たな規則を追加すればよい。

9.2 プログラム

```
1 and term() = match !tok with
2     L.ONE '(' -> (eat(L.ONE '(') ; term() ; eat(L.ONE ')'))
3     |L.VID s -> (eat(L.VID "") ; eat(L.IS) ; arithmexp())
4     | _ -> predicate()
```

9.3 プログラムの説明

問題 2 で作成した関数 term に次のプログラムを加える。先頭の文字が VID であれば VID を eat し、IS を eat、arithmexp() を実行する。

9.4 実行結果

このように正常に動作していることがわかる。

```
| tusedls00.ed.tus.ac.jp - Tera Term VT
ファイル(E) 編集(E) 設定(S) コントロール(Q)

^CInterrupted.
# open Parser;;
# run();;
?- KOTAE is 2.
VID(KOTAE)isNUM(2)ONE(.)
KAITOU is 15.
VID(KAITOU)isNUM(15)ONE(.)
SEISEKI is 9.
VID(SEISEKI)isNUM(9)ONE(.)
```