Tugas Pemrograman

IF2124 Teori Bahasa Formal dan Otomata HTML Checker dengan Pushdown Automata (PDA) Semester I Tahun 2023/2024

Disusun Oleh:

Ariel Herfrison (13522002)

Irfan Sidiq Permana (13522007)

Bryan Cornelius Lauwrence (13522033)



SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA INSTITUT TEKNOLOGI BANDUNG BANDUNG

2023

BAB 1 TEORI DASAR

Pushdown Automata adalah otomata yang mendefinisikan Context-Free Language. Pada intinya, Pushdown Automata adalah ϵ -NFA yang diperlengkapi dengan stack. Pushdown Automata dapat menerima input berdasarkan *accepting state* atau berdasarkan stack yang kosong. Salah satu subset PDA adalah deterministic PDA. Sebuah PDA adalah deterministic PDA jika dan hanya jika setiap $\delta(q, a, X)$ adalah kosong atau *singleton* dan jika $\delta(q, a, X)$ tidak kosong, maka $\delta(q, \epsilon, X)$ harus kosong. PDA dapat menerima kondisi akhir berdasarkan *stack* kosong atau *state* yang ditentukan sebagai *state* akhir (*final state*).

PDA dapat dimanfaatkan sebagai *bug finder*, salah satunya adalah *bug finder* untuk HTML. HTML (Hypertext Markup Language) adalah salah satu *markup language* untuk menyusun sebuah website dan kini digunakan oleh hampir semua website yang ada di internet. Sintaks HTML terdiri dari elemen tag yang dapat berupa elemen yang memiliki tag tutup, atau elemen yang tidak memiliki tag tutup (*void element*). Tiap elemen tag dapat memiliki suatu atribut global yang dapat dimiliki semua tag, misalnya *id*, *class*, *style*, dan sebagainya, serta atribut khusus yang hanya dimiliki oleh tag tertentu, misalnya *src*, *href*, dan sebagainya.

Untuk mengecek error atau tidaknya suatu file HTML, PDA sebagai HTML bug finder yang dibuat akan membaca elemen tag, lalu menyimpannya di dalam stack dan mengeluarkannya dari stack jika ditemukan tag penutupnya. Khusus void element, PDA akan mengeluarkan tag dari stack setelah menerima karakter '>' dari tag pembuka. PDA juga akan berpindah-pindah state bergantung posisi pengecekan saat ini. Pengecekan atribut suatu tag dilakukan dengan membaca tag buka terlebih dahulu lalu membaca atributnya jika ada. Kemudian, PDA akan mengecek atribut tersebut valid atau tidaknya atribut tersebut.

BAB 2 HASIL PDA

Berikut adalah format definisi PDA (PDA menerima *final state*)

```
Q P F # total states
a # input word symbols
Z Y # stack symbols
Q # starting state
Z # starting stack
F # accepting states
F # E - accepts with empty stack or F - accepts with accepting
state
Q a Z Q Y,Z # list of productions (current state, read from
word, take from stack, next state, add to stack)
QaYQY,Y
Q e Z P Z
Q e Y P Y
PaZPe
РаУРе
P e Z F e
```

Berikut adalah hasil definisi PDA HTML Checker:

```
00 H B F Text UT UAttr Dead
" < > / id=" class=" style=" <html </html> <head </head> <title
title> <script src=" script> <link href=" rel= <body </body>
<h1 h1 h1> <h2 h2 h2> <h3 h3 h3> <h4 h4 h4> <h5 h5 h5> <h6 h6
h6> <a a a> <imq imq alt=" src= <script script> <link <p p p>
<br br <hr hr <em em em> <b b r utton b> <abbr abbr>
<strong strong strong> <small small> <div div </div>
<button button type="submit" type="reset" type="button" button>
<form form form> <input input type="text" type="password"</pre>
type="email" type="number" type="checkbox" <table table
action=" method="POST" method="GET"  
 <!- !- -
ZO Ga GaT Html Head Body Title HScript BScript HLinkB HLinkA
BLinkB BLinkA H1 H2 H3 H4 H5 H6 P Br Em Bold Abbr Strong Small
Hr Div A ImgB ImgA Button Form FormN Input Table TR TD TH K
KOO KH KB KF KUT KA
Q0
7.0
F
F
```

```
<!- %
             Text KQ0,%
    <!- %
             Text KH, %
Η
В
    <!- %
             Text KB, %
   < % Dead
                 용
Dead !-
       용
             Text KF, %
UT !-
       용
            Text KUT, %
            Text KA
Text -
        K
             Text K
Text %
         K
Text -
      KA Text e
Text %
       KA
             Text K
Text >
        KQ0 Q0
Text >
         KΗ
             Η
                  е
Text >
      KB
             В
                  е
Text >
        KF
             F
Text >
         KUT Text e
      KQ0 Text K, KQ0
Text %
Text %
      KH Text K,KH
Text %
            Text K, KB
        KB
      KF
Text %
             Text K, KF
      KUT Text K, KUT
Text %
Text "
       Ga
             UAttr e
Text %
         Ga
             Text Ga
       GaT
                \mathtt{UT}
                          GaT
Text <
       FormN
Text <
                 UT
                          FormN
      < DivN
                 UT
Text
                       DivN
Text %
       GaT
                 Text GaT
Text %
                 Text FormN
         FormN
Text
       % DivN
                 Text DivN
UAttr
         id=" %
                 Text Ga, %
UAttr class=" % UAttr style=" %
        class=" % Text Ga,%
                       Text Ga, %
             Html Q0
UAttr
        >
                         Html
            Head H
UAttr
                          Head
        >
UAttr
         >
             Body B
                         Body
        /
UT
             GaT
                    UT
Q0
        <html Z0 UAttr
                          Html,Z0
Q0
    <head Html UAttr
                          Head, Html
            Head UAttr Title, Head
Н
    <title
UAttr > Title Text GaT, Title
UAttr > HScript Text GaT, HScript
UT title> Title H e
    <script Head UAttr</pre>
Н
                           HScript, Head
UAttr src=" HScript
                           Text Ga, HScript
UT
    script>
             HScript H
Η
    link
             Head UAttr
                           HLinkB, Head
UAttr href=" HLinkB Text Ga, HLinkB
```

```
UAttr rel= HLinkB UAttr HLinkA
UAttr " HLinkA Text Ga, HLinkA
UAttr href=" HLinkA Text Ga, HLinkA
       > HLinkA H
UAttr
                          е
H </head> Head H
                      е
   <body Html UAttr Body,Html</pre>
    <h1 Body UAttr H1,Body
h1 DivN UAttr H1,DivN
UT
UT h1 FormN UAttr H1, FormN
UAttr > H1 Text GaT, H1
  h1> H1 B e
UT
В
    <h2 Body UAttr H2, Body
UT h2 DivN UAttr H2, DivN
UT h2 FormN UAttr H2, FormN
UAttr
       > H2 Text GaT, H2
UT h2> H2 B e
    <h3 Body UAttr H3,Body
h3 DivN UAttr H3,DivN
В
UT
UT h3 FormN UAttr H3, FormN
UAttr > H3 Text GaT, H3
UT h3> H3 B e
    <h4 Body UAttr H4,Body
В
UT h4 DivN UAttr H4, DivN
UT h4 FormN UAttr H4, FormN
UAttr > H4 Text GaT, H4
UT h4> H4 B
                 е
B <h5 Body UAttr H5,Body
UT h5 DivN UAttr H5,DivN
UT h5 FormN UAttr H5, FormN
UAttr > H5 Text GaT, H5
UT h5> H5 B
                 е
   <h6 Body UAttr H6,Body
h6 DivN UAttr H6,DivN
В
UT
UT h6 FormN UAttr H6, FormN
UAttr > H6 Text GaT, H6
UT h6> H6 B e
    <a Body UAttr A,Body a DivN UAttr A,DivN
В
UT
UT a FormN UAttr A, For UAttr href=" A Text Ga, A UAttr > A Text GaT, A UT a> A B e
       FormN UAttr A, FormN
    <img Body UAttr ImgB,Body</pre>
UT img DivN UAttr ImgB, DivN
UT img FormN UAttr ImgB, FormN
UAttr alt=" ImgB Text Ga, ImgB
```

```
UAttr src= ImgB UAttr ImgA
UAttr " ImgA Text Ga, ImgA
UAttr alt=" ImgA Text Ga, ImgA
        > ImgA B e
UAttr
B <script Body UAttr BScript, Body
UT script DivN UAttr Bscript, DivN
UT script FormN UAttr BScript, FormN
UAttr src=" BScript Text Ga, BScript
UT script> BScript B e
B <link Body UAttr BLinkB, Body
UT link DivN UAttr BLinkB, DivN
UT link FormN UAttr BLinkB, FormN
UAttr href="BLinkB Text Ga, BLinkB
UAttr
        rel= BLinkB UAttr BLinkA
        " BLinkA Text Ga, BLinkA
UAttr
UAttr href=" BLinkA Text Ga, BLinkA UAttr > BLinkA B e
B <p Body UAttr P,Body UT p DivN UAttr P,DivN
UT p FormN UAttr P,FormN UAttr > P Text GaT,P UT p> P B e
B <br/>br Body UAttr Br,Body
UAttr > Br B
                        е
B e GaT Text GaT
B <hr Body UAttr Hr,Body
UT hr DivN UAttr Hr,DivN
UT hr FormN UAttr Hr, FormN
UAttr > Hr B e
UT em GaT UAttr Em, GaT
UT em DivN UAttr Em, DivN
UT em FormN UAttr Em, FormN
UAttr > Em Text GaT, Em
UT em> Em Text e
UT b GaT UAttr Bold, GaT
UAttr r Bold UAttr Br
UAttr utton Bold UAttr Button
UT b FormN UAttr Bold, FormN
UT b DivN UAttr Bold, DivN
UAttr > Bold Text GaT, Bold
UT b> Bold Text e
UT abbr GaT UAttr Abbr, GaT
UT abbr FormN UAttr Abbr, FormN UT abbr DivN UAttr Abbr, DivN
UAttr > Abbr Text GaT, Abbr UT abbr> Abbr Text e
```

```
UT strong GaT UAttr Strong, GaT
UT strong FormN UAttr Strong, FormN
UT strong DivN UAttr Strong, DivN
UAttr > Strong Text GaT, Strong UT strong> Strong Text e
UT small GaT UAttr Small, GaT
UT small FormN UAttr Small, FormN
UT small DivN UAttr Small, DivN
UAttr > Small Text GaT, Small
UT small> Small
                      Text e
В
    <div Body UAttr Div,Body</pre>
    div DivN UAttr Div, DivN
UT
UT div FormN UAttr Div, FormN
UT
    div
            DivN UAttr
                           Div, DivN
        > Div Text DivN, Div
UAttr
UT / DivN UT e
UT div> Div B
B e DivN Text DivN
B <button Body UAttr Button, Body
       > Button Text GaT, Button
UAttr
        type="submit" Button UAttr
        type="reset" Button UAttr
UAttr
                                         Button
UAttr type="button" Button UAttr Button
UT button> Button B e
В
   <form Body UAttr Form, Body</pre>
UT form DivN
                 UAttr Form, DivN
UT div FormN UAttr Div, FormN
UAttr method="GET" Form UAttr Form
         method="POST" Form UAttr
UAttr
                                   Form
UAttr action=" Form Text Ga, Form UAttr > Form Text FormN, Form
UT / FormN UT e
UT form> Form B e
B e FormN Text FormN

B <input Body UAttr Input,Body

UT input DivN UAttr Input,DivN
UT input FormN UAttr Input, FormN
UAttr type="text" Input UAttr Input
        type="password" Input UAttr Input
UAttr
        type="email" Input UAttr
UAttr
                                         Input
       type="number" Input UAttr Input
UAttr
        type="checkbox" Input UAttr Input
UAttr
UAttr > Input B
                           е
B <table Body UAttr Table, Body UT table FormN UAttr Table, FormN
UT table DivN UAttr Table, DivN
```

```
UAttr
               Table
                               Table
В
     <tr
          Table
                    UAttr
                               TR, Table
UAttr
                    В
          >
               TR
                          TR
В
     <th
          TR
               UAttr
                          TH, TR
В
     <td
          TR
                          TD, TR
               UAttr
                    Text GaT, TH
UAttr
          >
               TH
UAttr
          >
               TD
                    Text GaT, TD
UT
     th>
          TH
               В
UT
     td>
          TD
               В
                    е
В
     TR
                    B
                          е
В
     Table
                          В
                               е
В
     </body>
               Body B
                          е
В
     </html>
               Html F
```

Berikut adalah program PDA dalam python

```
mport os
import argparse
state list = []
input_list = []
stack_symbols = []
stack = []
final states = []
accept condition = None
transition_table = []
def printPda():
def inputAccepted(input, state, stack):
```

```
return temp
def epsilonTransition(state, stack):
    can epsilon = True
                    stack.pop()
                        stack.append(j)
dirname = os.path.dirname(__file__)
parser = argparse.ArgumentParser()
parser.add argument("a")
parser.add argument("b")
args = parser.parse args()
nama file pda = args.a
nama_file_pda = os.path.join(dirname,nama_file_pda)
file_pda = open(nama_file_pda,"r")
nama file html = args.b
nama file html = os.path.join(dirname, nama file html)
file html = open(nama file html,"r")
   state_list.append(word)
for word in file pda.readline().split():
    input list.append(word)
for word in file pda.readline().split():
    stack symbols.append(word)
state = file pda.readline().split()[0]
for word in file pda.readline().split():
   stack.append(word)
for word in file_pda.readline().split():
    final_states.append(word)
accept condition = file pda.readline().split()[0]
for line in file_pda:
    if (line.strip()):
```

```
Lines = file_html.readlines()
word_buffer = ""
m=1
last input = m
last_input_char = 0
                        stack.pop()
                                    stack.append(temp)
                                   stack.append(j)
```

```
if i[0] == state and i[1] == '%' and (stack[-1] == i[2] or i[2] == '%'):
                    stack.pop()
                                stack.append(temp)
                                stack.append(j)
epsilonTransition(state,stack)
else:
            temp.append(i[1])
```

BAB 3 IMPLEMENTASI DAN PENGUJIAN

1. Spesifikasi Teknis Program

Program dijalankan dalam command line dengan memanggil program beserta 2 argumen. Argumen pertama berupa text file (.txt) berisi definisi PDA. Argumen kedua berupa html file (.html) berisi kode HTML. Program kemudian akan membaca text file dan mencatat list states, list input word symbols, list stack symbols, starting state, starting stack, accepting states, accept condition, dan list of productions. Kemudian, program akan membaca html file dan menjalankan algoritma PDA untuk mencari tahu apakah html diterima atau tidak. Jika tidak diterima, program akan menunjukkan line di mana kode tidak sesuai dan menunjukkan expected input yang seharusnya diberikan. Pada program html checker kami, program akan melakukan epsilon transition terlebih dahulu sebelum lanjut memroses input (karakter pada file html).

2. Hasil Pengujian

1. Test Case 1

Expected Output: **Rejected** (head muncul setelah body)

Output:

```
PS C:\Ariel\Kuliah\MatKul\Tingkat 2\TBFO\Tubes\Tubes-IF2124-TBFO-2023-2024\src> python pda_machine.py PDA.txt tes_html.html Syntax Error
Terjadi kesalahan ekspresi pada line 1 : '<html>' karakter ke-7 : ''
Expected input: <head
```

2. Expected Output: **Rejected** (tag pertama bukan html)

Output:

```
PS C:\Ariel\Kuliah\MatKul\Tingkat 2\TBFO\Tubes\Tubes-IF2124-TBFO-2023-2024\src> python pda_machine.py PDA.txt tes_html.html
Syntax Error
Terjadi kesalahan ekspresi pada line 1 : '<hmif>' karakter ke-1 : '<'
Expected input: <html
```

3. Expected Output: **Rejected** (tidak punya head)

Output:

```
PS C:\Ariel\Kuliah\MatKul\Tingkat 2\TBF0\Tubes\Tubes-IF2124-TBF0-2023-2024\src> python pda_machine.py PDA.txt tes_html.html Syntax Error
Terjadi kesalahan ekspresi pada line 1 : '<html>' karakter ke-7 : ''
Expected input: <head
```

4. Accepted

Output:

PS C:\Ariel\Kuliah\MatKul\Tingkat 2\TBFO\Tubes\Tubes-IF2124-TBFO-2023-2024\src> python pda_machine.py PDA.txt tes_html.html Accepted

5. Expected Output: Accepted

Output:

6. Expected Output: **Rejected** (Atribut wajib src pada img tidak ada)

Output:

```
PS C:\Ariel\Kuliah\MatKul\Tingkat 2\TBFO\Tubes\Tubes-IF2124-TBFO-2023-2024\src> python pda_machine.py PDA.txt tes_html.html Syntax Error
Terjadi kesalahan ekspresi pada line 9 : ' <img alt="Welcome Banner">' karakter ke-30 : '>'
Expected input: src=
```

7. Expected Output: Accepted

```
<html>
  <title>Simple Webpage</title>
</head>
<body>
<h2>HTML Forms</h2>
<form action="/action page.php" method="POST">
 <h5 class="label">First name:</h5><br>
 <input type="text" id="fname"><br>
 <h5 class="label">Last name:</h5><br>
 <input type="text" id="lname"><br><br>
 <button type="submit">Submit
</form>
If you click the "Submit" button, the form-data will be sent to a
page called "/action page.php".
</body>
</html>
```

Output:

8. Expected Output: **Rejected** (tidak termasuk value yang diperbolehkan (POST, GET))

```
<html>
<head>
  <title>Simple Webpage</title>
</head>
<body>
<h2>HTML Forms</h2>
<form action="/action page.php" method="TEMBAK">
 <div id="label">First name:</div><br>
 <input type="text" id="fname"><br>
  <div id="label">Last name:</div><br>
 <input type="text" id="lname"><br><br>
 <button type="submit">Submit
</form>
If you click the "Submit" button, the form-data will be sent to a
page called "/action page.php".
</body>
</html>
```

Output:

```
PS C:\Ariel\Kuliah\MatKul\Tingkat 2\TBFO\Tubes\Tubes-IF2124-TBFO-2023-2024\src> python pda_machine.py PDA.txt tes_html.html Syntax Error
Terjadi kesalahan ekspresi pada line 10 : '<form action="/action_page.php"_method="TEMBAK">' karakter ke-32 : ''
Expected input: method="GET" || method="POST" || action=" || >
```

9. Expected Output: Accepted

Output:

PS C:\Ariel\Kuliah\MatKul\Tingkat 2\TBFO\Tubes\Tubes-IF2124-TBFO-2023-2024\src> python pda_machine.py PDA.txt tes_html.html

10. Expected Output: **Rejected** (bukan void element, wajib ditutup)

```
<html>
<head>
    <title>Simple Webpage</title>
        <script>
            document.getElementById("demo").innerHTML = "Hello JavaScript!";
            </script>
            </head>
            <body>
            <hl>The script element</hl>
             </body>
            </body>
            </html>
```

Output:

```
PS C:\Ariel\Kuliah\MatKul\Tingkat 2\TBF0\Tubes\Tubes-IF2124-TBFO-2023-2024\src> python pda_machine.py PDA.txt tes_html.html Syntax Error
Terjadi kesalahan ekspresi pada line 14 : '</body>' karakter ke-3 : 'b'
Expected input: p>
```

11. Expected Output: Accepted

Output:

PS C:\Ariel\Kuliah\MatKul\Tingkat 2\TBF0\Tubes\Tubes-IF2124-TBF0-2023-2024\src> python pda_machine.py PDA.txt tes_html.html Accepted

REFERENSI

Link repository github: github.com/BryanLauw/Tubes-IF2124-TBFO-2023-2024

Link diagram state: https://tinyurl.com/37eer7u4

Pembagian Tugas:

NIM	NAMA	PEMBAGIAN TUGAS
13522002	Ariel Herfrison	Membuat fungsi PDA dalam python
13522007	Irfan Sidiq P	Merumuskan definisi PDA
13522033	Bryan Cornelius	Merumuskan definisi PDA