

Αρχές Γλωσσών και Μεταφραστών Εργαστηριακή 'Ασκηση Ακαδημαικό Έτος 2018-2019

Μέλη Ομάδας:

1)Τουρλίδα Βάγια **ΑΜ:**6233

2)Γεωργάτος Βαγγέλης ΑΜ:6015

3)Τσιούμα Αγγελική **ΑΜ:**6246

Γενική Θεωρία

Το Bison είναι ένας γενικού σκοπού αναλυτής ο οποίος μετατρέπει μια contex-free γλώσσα σε έναν LALR(1) ή έναν αναλυτή GLR για αυτή τη γραμματική.

Ένας αναλυτής LR είναι ουσιαστικά ένας αλγόριθμος shift-reduce (bottom-up) που καθοδηγείται από πίνακες ανάλυσης και μια στοίβα.

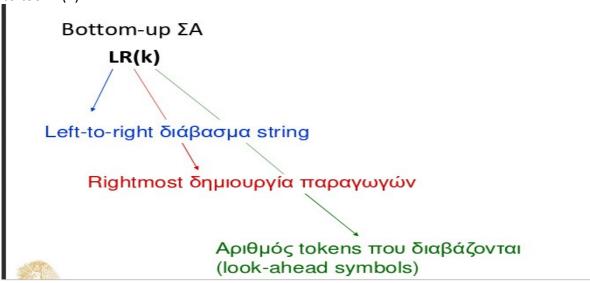
Αυτοί κατασκευάζουν το δέντρο συντακτικής ανάλυσης ξεκινώντας απο τα φύλλα και προχωρώντας προς τη ρίζα αναγνωρίζοντας πότε μια συλλογή φύλλων ή άλλων κόμβων μπορεί να συνενωθούν ως παιδιά ενός νέου κόμβου του δένδρου.

Οι αναλυτές αυτοί δε χρειάζεται να προβλέψουν τι θα δούν στην συνέχεια όπως κάνουν οι top-down αντ'αυτού μεταφέρουν τις λεκτικές μονάδες στο δάσος των κανόνων μέχρι να αναγνωριστεί στο δεξιό μέλος κάποιου συντακτικού κανόνα που αμέσως κάνει reduce στο αντίστοιχο αριστερό μέλος.

Ένας τέτοιος αναλυτής θυμάται τις καταστάσεις από τις οποίες έχει περάσει τοποθετώντας τες στη στοίβα μαζί με τα σύμβολα της γραμματικής.Με λίγα λόγια οι καταστάσεις είναι αυτές που καθοθηγούν τον αλγόριθμο της συντακτικής ανάλυσης και όχι τα σύμβολα.Παρατίθεται ένα παράδειγμα για να γίνει πιο κατανοητό το παρακάτω

π.χ Όταν ο συνδυασμός κατάστασης και εισόδου μας λέει ότι χρειάζεται να κάνουμε reduce χρησιμοποιώντας τον κανόνα Α->α αφαιρούμε τόσα σύμβολα από την στοίβα όσο και το μήκος του α και τα αντικαθιστούμε με το Α.

Το εργαλείο bison είναι γεννήτορας συντακτικών αναλυτών για γραμματικές χωρίς συμφραζόμενα τύπου LR(1)



Οι αναλυτές τύπου LR(1) μας βοηθούν ώστε να αποφύγουμε τις διάφορες συγκρόυσης τύπου είτε reduce-reduce είτε shift-reduce.

Μέρος Α

Ερώτημα 1

```
Η Μορφή BNF του συντακτικού ορισμού της γλώσσας παρατίθεται παρακάτω:
<root> ::= <Workbook> ;
<Workbook > ::= OPEN WORKBOOK CLOSE <Styles_rule> <Worksheet> OP_CLOSE WORKBOOK CLOSE;
<Worksheet> ::= OPEN WORKSHEET <Protected> NAME EQUAL QUOTE_STR <Protected> CLOSE <Table> OP_CLOSE
WORKSHEET CLOSE;
<Protected> ::= /*e*/ | PROTECTED EQUAL QUOTE BOOL;
<Styles_rule> ::= /*e*/ | <Styles_rule> OPEN STYLES CLOSE <Style> OP_CLOSE STYLES CLOSE ;
<Style> ::= /*e*/ | <Style> OPEN STYLE ID EQUAL QUOTE_STR CLOSE OP_CLOSE STYLE CLOSE <Style> ;
<Table> ::= /*e*/ | OPEN TABLE <ExpandedColumnCount> <ExpandedRowCount> StyleID CLOSE <Atr> OP_CLOSE
TABLE CLOSE;
<ExpandedColumnCount> ::= /*e*/ | EXPANDEDCOLUMN EQUAL QUOTE_INT ;
<ExpandedRowCount>::=/*e*/ | EXPANDEDROW EQUAL QUOTE INT;
<StyleID> ::= /*e*/ | STYLEID EQUAL QUOTE_STR;
<Atr> ::= <Column> <Row> ;
<Column>::= /*e*/ | <Column> OPEN COLUMN <Hidden> <Width> <StyleID> CL CLOSE;
<Hidden>::= /*e*/ | HIDDEN EQUAL QUOTE BOOL;
<Width>::=/*e*/ | WIDTH EQUAL QUOTE_INT;
<Row> ::= /*e*/ | OPEN ROW <Height> <Hidden> <StyleID> CLOSE <Cell> OP_CLOSE ROW CLOSE <Row> ;
<Height> ::= /*e*/ | HEIGHT EQUAL QUOTE_INT;
<Cell> ::= /*e*/ | <Cell> OPEN CELL <MergeAcross> <MergeDown> <StyleID> CLOSE <Data> OP_CLOSE CELL CLOSE;
<MergeAcross> ::= /*e*/ | MERGEACROSS EQUAL QUOTE_INT ;
<MergeDown> ::= /*e*/ | MERGEDOWN EQUAL QUOTE_INT;
<Data> ::= /*e*/ | <Data> OPEN DATA TYPE EQUAL <Type_Cont> CLOSE <Content> OP_CLOSE DATA CLOSE ;
<Content> ::= /*e*/ | <Content> STRING;
<Type Cont> ::= DATE
 | QUOTE_BOOL
 | QUOTE_STR
 | QUOTE_INT
```

Ερωτημα 2

Σχετικά με τους ελέγχους που απαιτούνται στο ερώτημα 2 έγινε χρήση κώδικα σε γλώσσα C και της δυνατότητας των semantic actions που προσφέρουν τα εργαλεία bison/flex.

Ποιο συγκεκριμένα , στο ερώτημα 2i κάθε φορά που συναντάμε τον κανόνα που παράγει το ss:Style αποθηκεύουμε τις τιμές ID σε ένα double char array και εν συνεχεία γίνεται ο ζητούμενος έλεγχος στην main.Την ίδια διαδικασία συναντάμε και στο ερώτημα 2ii με την διαφορά ότι εδώ οι παραπάνω ενέργειες εφαρμόζονται στον κανόνα που παράγει το ss:StyleID. Στα ερωτήματα, (2iii,2iv) ακολουθείται παρόμοια τεχνική μόνο που εδώ απλά αποθηκεύουμε την τιμή που δίνεται στους κανόνες ExpandedColumnCount ή ExpandedRowCount με την τιμή ενός μετρητή ο οποίος αυξάνεται κατά ένα κάθε φορά που συναντά κανόνα column ή row/

Τα παράπανω ερωτήματα υλοποιήθηκαν με τον κώδικα που παρατίθεται παρακάτω:

Αρχείο Bison

```
my_parser.y
```

```
%{
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#define YYERROR VERBOSE
extern int yylineno;
        extern char* yytext;
        extern FILE *yyin;
        extern FILE *yyout;
char **sids; // style ids decleared in the header
char **ids; // style ids in the rest of the text
int errors = 0;
int ECCvalue = 0;
int ERCvalue = 0;
int rcount = 0;
int ccount = 0;
int c1 = 0;
int c2 = 0:
%}
%error-verbose
%union {
        int int t;
        char *str t;
        int bool t;
        }
%token WORKBOOK DATA STYLES STYLE WORKSHEET TABLE TOKROW COLUMN CELL
%token ROW
%token ID PROTECTED NAME EXPANDEDROW EXPANDEDCOLUMN STYLEID WIDTH HIDDEN HEIGHT MERGEACROSS
MERGEDOWN TYPE
%token BOOL STRING SIGNED_INT DATE STR NUMBER QUOTE_STR QUOTE_INT QUOTE_BOOL
%left CLOSE CL CLOSE
%right EQUAL OPEN OP CLOSE ROW
%type <int_t> QUOTE_INT
```

```
%type <str t> QUOTE STR
%%
root: Workbook;
Workbook: OPEN WORKBOOK CLOSE Styles rule Worksheet OP CLOSE WORKBOOK CLOSE;
Worksheet: OPEN WORKSHEET Protected NAME EQUAL QUOTE STR Protected CLOSE Table OP CLOSE WORKSHEET
Protected: /*e*/ | PROTECTED EQUAL QUOTE BOOL;/*dwstou timh boolean*/
Styles rule: /*e*/ | Styles rule OPEN STYLES CLOSE Style OP CLOSE STYLES CLOSE;
Style: /*e*/ | Style OPEN STYLE ID EQUAL QUOTE_STR CLOSE OP_CLOSE STYLE CLOSE Style {sids =
(char^*) realloc(sids,1); sids[c1] = (char^*) malloc(strlen($6) + 1); strcpy(sids[c1],$6); c1++;};
Table:/*e*/ | OPEN TABLE ExpandedColumnCount ExpandedRowCount StyleID CLOSE Atr OP_CLOSE TABLE CLOSE;
ExpandedColumnCount: /*e*/ | EXPANDEDCOLUMN EQUAL QUOTE INT {ECCvalue = $3;};
ExpandedRowCount: /*e*/ | EXPANDEDROW EQUAL QUOTE_INT {ERCvalue = $3;};
StyleID: /*e*/ | STYLEID EQUAL QUOTE STR {ids = (char**)realloc(ids,1); ids[c2] = (char*)malloc(strlen($3) + 1);
strcpy(ids[c2],$3);c2++;};
Atr: Column Row
Column: /*e*/ | Column OPEN COLUMN Hidden Width StyleID CL_CLOSE {ccount ++; } ;
Hidden: /*e*/ | HIDDEN EQUAL QUOTE_BOOL;
Width: /*e*/ | WIDTH EQUAL QUOTE INT;
Row: /*e*/ | OPEN ROW Height Hidden StyleID CLOSE Cell OP CLOSE ROW CLOSE Row {rcount ++; };
Height: /*e*/ | HEIGHT EQUAL QUOTE INT;
Cell: /*e*/ | Cell OPEN CELL MergeAcross MergeDown StyleID CLOSE Data OP_CLOSE CELL CLOSE;
MergeAcross: /*e*/ | MERGEACROSS EQUAL QUOTE INT;
MergeDown: /*e*/ | MERGEDOWN EQUAL QUOTE INT;
Data: /*e*/ | Data OPEN DATA TYPE EQUAL Type_Cont CLOSE Content OP_CLOSE DATA CLOSE;
Content: /*e*/ | STRING Content;
Type Cont: QUOTE BOOL
  | QUOTE STR
  | QUOTE_INT
%%
int yywrap(void)
 return 0;
void yyerror(char *s){
    errors++;
        printf("\n----- ERROR AT LINE #%d. %s \n\n", yylineno,s);
}
int main (int argc, char **argv){
        argv++;
        argc--;
        int i,j,l;
        char *p;
```

```
char *q;
         int found[c1];
         int found2[c2];
         errors=0;
          int errflag1 = 0;
         int errflag2 = 0;
         int errflag3 = 0;
         if(argc>0)
                  yyin=fopen(argv[0], "r");
         else
                  yyin=stdin;
         yyparse();
         if (ERCvalue != rcount){
         printf("Beware that the decleared rows are not equal to the existing rows\n");
         errflag1 = 1;
}
         if (ECCvalue != ccount){
         printf("Beware that the decleared columns are not equal to the existing columns\n");
         errflag1 =1;
                                     }
         for (i=0;i<c1;i++){
                  found[i] = 0;
                                     }
         for (i=0;i<c1;i++){}
                  for (j=0;j<c1;j++){
                  if( strcmp(sids[i],sids[j]) == 0){
                            found[i]++;
                  }
                  }
                  }
                  for (i=0;i<c1;i++){
                  if (found[i] > 1){
                            errflag2 = 1;
                                     break;
                                     }
                                     }
                  for (i=0;i<c1;i++){}
                  found[i] = 0;
                                     }
         for (i=0;i<c2;i++){}
                  for (j=0;j<c1;j++){
                  if (strcmp(ids[i],sids[j]) == 0)
                            found2[i]++;
                            }
                            }
         for (i=0;i<c2;i++){
                  if (found2[i] == 0){
                            errflag3 =1;
                            break;
                                     }
```

```
if((errors==0) && (errflag1 == 0) && (errflag2 == 0)&&(errflag3 == 0)) {
    printf("------ PARSING COMPLETED, VALID PROGRAM.\n\n");
}

if (errflag1 == 1){
    printf("ERROR AT YOUR ROWS/COLUMNS DECLERATIONS\n");
}

if (errflag2 == 1){
    printf("DUPLICATE STYLE ID DECLERATION\n");
    }

if (errflag3 == 1){
    printf("STYLE ID NOT DECLARED ABOVE\n");
}
```

Αρχείο flex

XML_syntax_4.l

```
%{
        #include "y.tab.h"
        #include <stdio.h>
        #include <stdlib.h>
        #include <string.h>
        char *temp;
        char *stemp;
%option noyywrap
%option yylineno
%x comments
str [a-zA-Z0-9_]*[a-zA-Z_][a-zA-Z0-9_]*
quote [\"]
digit [0-9]
signed_integer{digit}+
white_spaces [ \t\n]
%%
"<!--" BEGIN(comments);
<comments>{
   [^-\n]*
  "-"+[^->\n]*
   "-->" BEGIN(INITIAL);
}
```

```
"Workbook"
                         {printf("\t WORKOBOOK: %s\n", yytext); return (WORKBOOK);}
"Worksheet"
                         {printf("\t WORKSHEET: %s\n", yytext); return (WORKSHEET);}
"Styles"
                    {printf("\t STYLES: %s\n", yytext); return (STYLES);}
"Style"
                   {printf("\t STYLE: %s\n", yytext);return (STYLE);}
"Table"
                   {printf("\t TABLE: %s\n", yytext); return (TABLE);}
"ss:ExpandedRowCount"
                                     {printf("\t ExpandedRowCount: %s\n", yytext); return (EXPANDEDROW);}
"ss:ExpandedColumnCount"
                                        {printf("\t ExpandedColumnCount: %s\n", yytext); return
(EXPANDEDCOLUMN);}
"ss:StyleID"
                        {printf("\t : StyleID %s\n", yytext); return (STYLEID);}
"ss:Name"
                       {printf("\t Name: %s\n", yytext); return (NAME);}
"Column"
                      {printf("\t Column: %s\n", yytext); return (COLUMN);}
"Row"
                   {printf("\t Row: %s\n", yytext); return (ROW);}
"Cell"
                  {printf("\t Cell: %s\n", yytext); return (CELL);}
"Data"
                   {printf("\t Data: %s\n", yytext); return (DATA);}
"ss:ID"
                   {printf("\t ID: %s\n", yytext); return (ID);}
                       {printf("\t Height: %s\n", yytext); return (HEIGHT);}
"ss:Height"
"ss:Hidden"
                        {printf("\t Hidden: %s\n", yytext); return (HIDDEN);}
                              {printf("\t MergeAcross: %s\n", yytext); return (MERGEACROSS);}
"ss:MergeAcross"
"ss:Width"
                       {printf("\t Width: %s\n", yytext); return (WIDTH);}
"ss:MergeDown"
                             {printf("\t MergeDown: %s\n", yytext); return (MERGEDOWN);}
"ss:Type"
                      {printf("\t Type: %s\n", yytext); return (TYPE);}
"ss:Protected" {printf("\t Protected: %s\n", yytext); return (PROTECTED);}
"\"Boolean\"" {printf("\t Boolean: %s\n", yytext); return (QUOTE_BOOL);}
{white_spaces} {}
          {printf("\t String: %s\n", yytext); return (STRING);}
{quote}{str}{quote} {printf("\t Quoted String: %s\n", yytext); yytext++; yytext[strlen(yytext)-1] = 0; yylval.str_t =
yytext; return (QUOTE STR);}
{quote}{signed_integer}{quote} {printf("\t Quoted signed_integer: %s\n", yytext); temp = yytext; temp++;
temp[strlen(temp)-1] = 0; yylval.int_t = atoi(temp); return (QUOTE_INT);}
"=" { printf("\t Equal: %s\n", yytext); return (EQUAL);}
"<ss:" {printf("\t Open %s\n", yytext); return (OPEN);}
">" {printf("\t Close %s\n", yytext); return (CLOSE);}
"</ss:" {printf("\t OP_Close %s\n", yytext); return (OP_CLOSE);}
"/>" {printf("\t CL Close %s\n", yytext); return (CL CLOSE);}
```

Για διευκόλινση μας ώστε να γίνεται πιο γρήγορα compile δημιουργήθηκε ένα αρχείο makefile.

Αρχείο makefile

default:

```
bison -d -y my_parser.y
flex XML_syntax_4.l
gcc -c y.tab.c lex.yy.c
gcc y.tab.o lex.yy.o -o parser
```

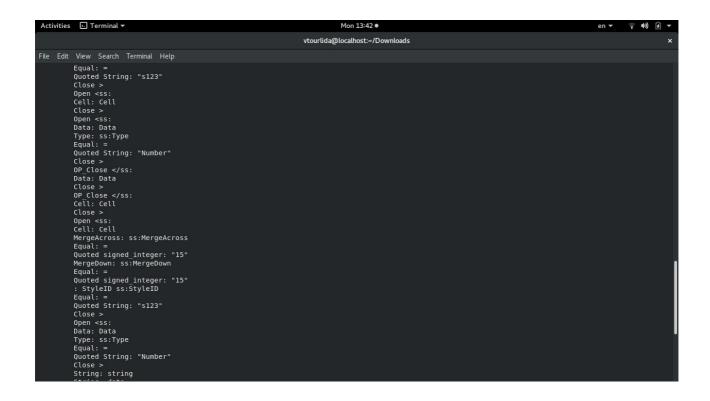
Για εισοδο το αρχείο **input.txt** :

```
<!-- Sxolio -->
<ss:Workbook>
<ss:Styles>
<ss:Style ss:ID="s123"></ss:Style>
<ss:Style ss:ID="s133"></ss:Style>
</ss:Styles>
</ss:Styles>
<ss:Worksheet ss:Name="sheet1" ss:Protected="Boolean">
<ss:Table ss:ExpandedColumnCount="2" ss:ExpandedRowCount="1" >
<ss:Column ss:Hidden="Boolean" ss:Width="25" ss:StyleID="s123" />
```

```
<ss:Column ss:StyleID="s123"/>
<ss:Row ss:Height="25" ss:Hidden="Boolean" ss:StyleID="s123">
<ss:Cell>
<ss:Data ss:Type="Number">
</ss:Data>
</ss:Cell>
<ss:Cell ss:MergeAcross="15" ss:MergeDown="15" ss:StyleID="s123" >
<ss:Data ss:Type="Number"> string data
</ss:Data>
</ss:Data>
</ss:Cell>
</ss:Row>
</ss:Table>
</ss:Workbook>
```

Screenshots αποτελεσμάτων:

```
vtourlida@localhost:~/Downloads
                                                                                                           Open <ss:
WORKOBOOK: Workbook
Close >
Open <ss:
STYLES: Styles
Close >
Open <ss:
STYLES: Style
ID: ss:ID
Equal: =
Quoted String: "s123"
Close >
OP Close </ss:
STYLE: Style
Close >
Open <ss:
STYLE: Style
ID: ss:ID
Equal: =
Quoted String: "s133"
Close >
OP Close </ss:
STYLE: Style
ID: ss:ID
Equal: =
Quoted String: "s133"
Close >
OP Close </ss:
STYLE: Style
Close >
OP Close </ss:
STYLES: Style
Close >
OP Close </ss:
STYLES: Styles
Close >
ITYLES: Styles
Close >
ITYLES: Styles
Close >
ITYLES: Styles
Close >
ITYLES: Styles
ITYLES: Styles
ITYLES: Styles
ITYLES: Styles
ITYLES: ITYL
                                                                                                                      Open <ss:
WORKOBOOK: Workbook
Activities ☑ Terminal ▼
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   Mon 13:41 ●
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 en ▼ 🛜 🕪 📝 ▼
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 vtourlida@localhost:~/Downloads
                                                                                                                   Open <ss:
TABLE: Table
ExpandedColumnCount: ss:ExpandedColumnCount
                                                                                                               ExpandedColumnCount: ss:ExpandedColumn
Equal: =
Quoted signed integer: "2"
ExpandedRowCount: ss:ExpandedRowCount
Equal: =
Quoted signed_integer: "1"
Close >
Open <ss:
Column: Column
Hidden: ss:Hidden
Equal: =
Boolean: "Boolean"
Width: ss:Width
Equal: =
Quoted signed_integer: "25"
: styleID ss:StyleID
Equal: =
                                                                                                               Gutted asymetation and the state of the stat
                                                                                                                      Equal: =
Boolean: "Boolean"
: StyleID ss:StyleID
```



Για είσοδο το **input2.txt** που περιέχει συντακτικα λάθη δηλ.

<!-- Sxolio --> <ss:Workbook>

δεν κλείνει ποτέ το Workbook το αποτέλεσμα είναι το εξής:

Screenshots αποτελεσμάτων:

```
ida@laptop-5lleli4n Downloads]$ ./parser input2.txt
Open <ss:
WORKOBOOK: Workbook
Close >
ERROR AT LINE #3. syntax error, unexpected $end, expecting OPEN
```

Sxolio >
<ss:workbook></ss:workbook>
<ss:styles></ss:styles>
<ss:style ss:id="s123"></ss:style>
<ss:style ss:id="s133"></ss:style>
<ss:worksheet ss:name="sheet1" ss:protected="Boolean"></ss:worksheet>
Screenshots αποτελεσμάτων:

ερώτημα 2) Για τον έλεγχο ορθότητας του ερωτήματος 2 σαν input χρησιμοποιήθηκε τα αρχεία dupStyleID.txt , allwrong.txt , wrongrowcount.txt , wrongcolumncount.txt , notdeclaredid.txt

Παρακάτω φαίνονται τα screenshots των αποτελεσμάτων

