

# Tým xhricma00 varianta vv-BVS

**Marek Hric**  
xhricma00

Mikuláš Lešiga  
xlesigm00

Roman Andraščík  
xandrar00

Adam Veselý  
xvesela00

December 5, 2024

## Rozdelenie bodov

xhricma00: 25%

xlesigm00: 25%

xandrar00: 25%

xvesela00: 25%

## Rozšírenia

ORELSE

UNREACHABLE

BOOLTHEN

FOR

WHILE

FUNEXP

## Rozdelenie prace :

---

Marek Hric :

- Lexikálna analýza
- Generovanie kodu

Mikuláš Lešiga :

- Syntaktická analýza bez výrazov
- Semantická analýza bez výrazov
- LL Gramatika
- LL tabuľka, precedenčná tabuľka

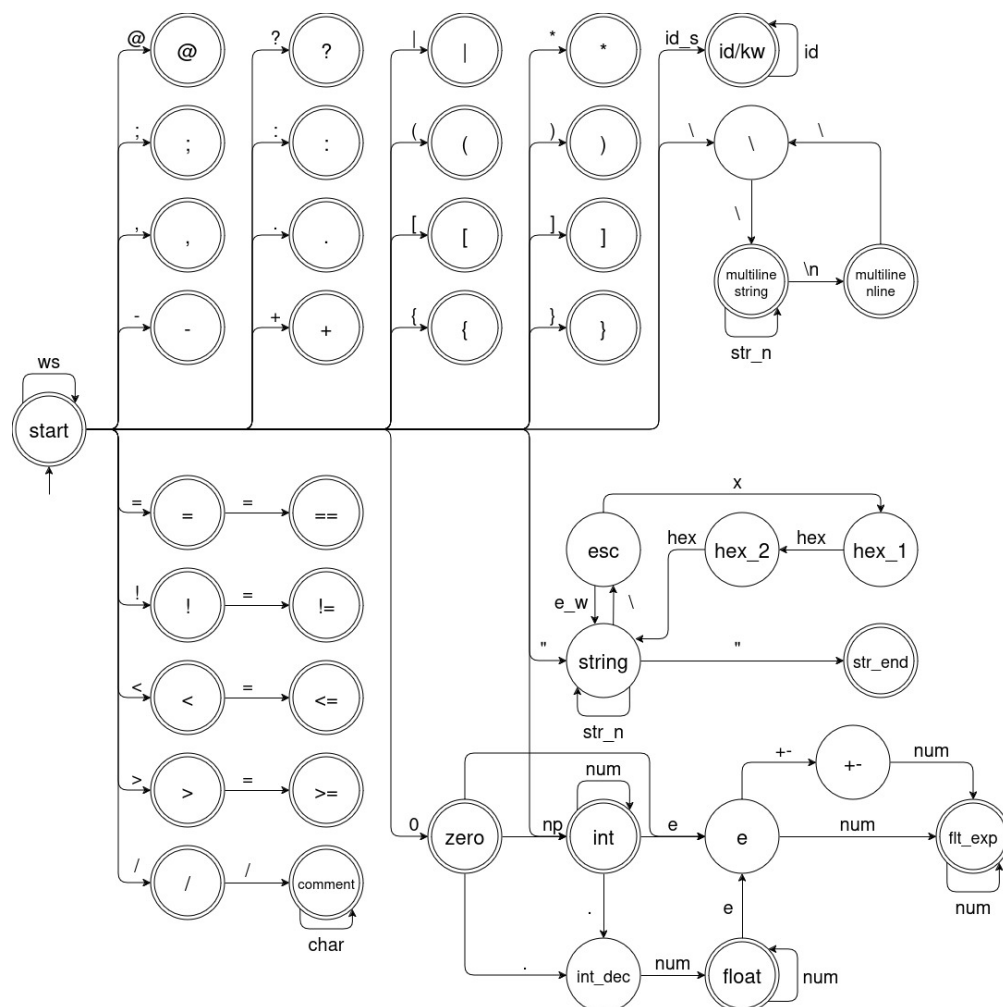
Roman Andraščík :

- Testy
- Pomocné funkcie
- Symtable
- Dokumentácia

Adam Veselý :

- Syntaktická analýza - výrazy
- Semantická analýza - výrazy
- Dátové štruktúry

## Diagram konečného automatu :



## Legenda:

- **ws** - white space
- **id\_s** - znaky začiatku identifikátora (`_a-zA-Z`)
- **id** - znaky identifikátora (`_a-zA-Z0-9`)
- **num** - čísla (`0-9`)
- **np** - čísla bez nuly (`1-9`)
- **hex** - hexa čísla (`0-9a-fA-F`)
- **e** - exponent (`eE`)
- **char** - ľubovoľný znak
- **e\_w** - znaky escape sekvencii bez `x` (`n,t,r`)
- **str\_n** - ľubovoľný znak bez `\n`

## LL-gramatika :

---

1.  $\langle prog \rangle \rightarrow \langle prolog \rangle \langle function\_def \rangle \langle end \rangle$
2.  $\langle prolog \rangle \rightarrow \text{const ID} = @import(\langle expression \rangle);$
3.  $\langle end \rangle \rightarrow EOF$
4.  $\langle function\_def \rangle \rightarrow \text{pub fn ID}(\langle param\_list \rangle) \langle function\_type \rangle \langle block \rangle \langle function\_def \rangle$
5.  $\langle function\_def \rangle \rightarrow \varepsilon$
6.  $\langle param\_list \rangle \rightarrow ID : \langle type\_complete \rangle \langle comma\_par\_found \rangle$
7.  $\langle param\_list \rangle \rightarrow \varepsilon$
8.  $\langle comma\_par\_found \rangle \rightarrow , \langle param\_list \rangle$
9.  $\langle comma\_par\_found \rangle \rightarrow \varepsilon$
10.  $\langle block \rangle \rightarrow \{ \langle statement \rangle \}$
11.  $\langle statement \rangle \rightarrow \langle var\_declaration \rangle \langle statement \rangle$
12.  $\langle statement \rangle \rightarrow ID \langle ID\_found \rangle \langle statement \rangle$
13.  $\langle statement \rangle \rightarrow \langle if\_statement \rangle \langle statement \rangle$
14.  $\langle statement \rangle \rightarrow \langle for\_loop \rangle \langle statement \rangle$
15.  $\langle statement \rangle \rightarrow \langle while\_loop \rangle \langle statement \rangle$
16.  $\langle statement \rangle \rightarrow \langle return\_statement \rangle \langle statement \rangle$
17.  $\langle statement \rangle \rightarrow \langle break \rangle \langle statement \rangle$
18.  $\langle statement \rangle \rightarrow \langle continue \rangle \langle statement \rangle$
19.  $\langle statement \rangle \rightarrow \varepsilon$
20.  $\langle ID\_found \rangle \rightarrow = \langle asgn\_found \rangle ;$
21.  $\langle ID\_found \rangle \rightarrow ( \langle expression\_list \rangle );$
22.  $\langle ID\_found \rangle \rightarrow : \langle while\_loop \rangle$
23.  $\langle var\_declaration \rangle \rightarrow \text{const ID} : \langle type\_complete \rangle = \langle asgn\_found \rangle ;$
24.  $\langle var\_declaration \rangle \rightarrow \text{var ID} : \langle type\_complete \rangle = \langle asgn\_found \rangle ;$
25.  $\langle if\_statement \rangle \rightarrow \text{if}(\langle expression \rangle) \langle if\_found \rangle$

26.  $\langle \text{if\_found} \rangle \rightarrow \langle \text{optional\_value} \rangle \langle \text{block} \rangle \langle \text{else\_statement} \rangle$
27.  $\langle \text{else\_statement} \rangle \rightarrow \text{else} \langle \text{block} \rangle$
28.  $\langle \text{else\_statement} \rangle \rightarrow \varepsilon$
29.  $\langle \text{optional\_value} \rangle \rightarrow |ID|$
30.  $\langle \text{optional\_value} \rangle \rightarrow \varepsilon$
31.  $\langle \text{while\_loop} \rangle \rightarrow \text{while}(\langle \text{expression} \rangle) \langle \text{optional\_value} \rangle \langle \text{optional\_statements} \rangle \langle \text{block} \rangle \langle \text{else\_statement} \rangle$
32.  $\langle \text{return\_statement} \rangle \rightarrow \text{return} \langle \text{expression} \rangle;$
33.  $\langle \text{expression\_list} \rangle \rightarrow \langle \text{expression} \rangle \langle \text{comma\_expr\_found} \rangle$
34.  $\langle \text{expression\_list} \rangle \rightarrow \varepsilon$
35.  $\langle \text{comma\_expr\_found} \rangle \rightarrow , \langle \text{expression\_list} \rangle$
36.  $\langle \text{comma\_expr\_found} \rangle \rightarrow \varepsilon$
37.  $\langle \text{type} \rangle \rightarrow i32$
38.  $\langle \text{type} \rangle \rightarrow f64$
39.  $\langle \text{type} \rangle \rightarrow []u8$
40.  $\langle \text{type} \rangle \rightarrow \text{bool}$
41.  $\langle \text{for\_loop} \rangle \rightarrow \text{for}(\langle \text{expression} \rangle) \langle \text{optional\_value} \rangle \langle \text{block} \rangle$
42.  $\langle \text{optional\_statements} \rangle \rightarrow \text{:} \langle \text{block} \rangle$
43.  $\langle \text{optional\_statements} \rangle \rightarrow \varepsilon$
44.  $\langle \text{type\_complete} \rangle \rightarrow \langle \text{question\_mark} \rangle \langle \text{type} \rangle$
45.  $\langle \text{question\_mark} \rangle \rightarrow ?$
46.  $\langle \text{question\_mark} \rangle \rightarrow \varepsilon$
47.  $\langle \text{single\_statement} \rangle \rightarrow \langle \text{var\_declaration} \rangle$
48.  $\langle \text{single\_statement} \rangle \rightarrow ID \langle ID\_found \rangle$
49.  $\langle \text{single\_statement} \rangle \rightarrow \langle \text{if\_statement} \rangle$
50.  $\langle \text{single\_statement} \rangle \rightarrow \langle \text{for\_loop} \rangle$
51.  $\langle \text{single\_statement} \rangle \rightarrow \langle \text{while\_loop} \rangle$

52.  $\langle single\_statement \rangle \rightarrow \langle return\_statement \rangle$
53.  $\langle single\_statement \rangle \rightarrow continue;$
54.  $\langle single\_statement \rangle \rightarrow break;$
55.  $\langle function\_type \rangle \rightarrow \langle type \rangle$
56.  $\langle function\_type \rangle \rightarrow void$
57.  $\langle asgn\_found \rangle \rightarrow \langle expression \rangle$
58.  $\langle continue \rangle \rightarrow continue;$
59.  $\langle break \rangle \rightarrow break \langle cycle\_current\_control \rangle$
60.  $\langle break \rangle \rightarrow continue \langle cycle\_current\_control \rangle$
61.  $\langle cycle\_current\_control \rangle \rightarrow ID$
62.  $\langle cycle\_current\_control \rangle \rightarrow \varepsilon$

## Precendecna-tabulka :

	.?	!	*/	+-	orelse	r	and	or	(	)	i	\$
.?	<								<	>	<	>
!		<	>	>	>	>	>	>	<	>	>	>
*/		<	>	>	>	>	>	>	<	>	<	>
+-		<	<	>	>	>	>	>	<	>	<	>
orelse		<	<	<	>	>	>	>	<	>	<	>
r		<	<	<	<		>	>	<	>	<	>
and		<	<	<	<	<	>	>	<	>	<	>
or		<	<	<	<	<	<	>	<	>	<	>
(	<	<	<	<	<	<	<	<	<	=	<	
)	>	>	>	>	>	>	>	>		>		>
i	>		>	>	>	>	>	>		>		>
\$	<	<	<	<	<	<	<	<	<		<	:)

LL-tabulka :

	const	00	=	import	exportations	(	)	+	-	EGP	pub	fin	e	:	.	+	(	)	-	var	#	else		while	return	32	64	(	)	us	void	bool	for	7	break	continue	
<ptrdiff_t>	1																																				
<size_t>	2									3	4	5	6																								
<ptrdiff_t>													7	8																							
<ptrdiff_t>													9	10																							
<ptrdiff_t>													11	12																							
<ptrdiff_t>													13	14																							
<ptrdiff_t>													15	16																							
<ptrdiff_t>													17	18																							
<ptrdiff_t>													19	20																							
<ptrdiff_t>													21	22																							
<ptrdiff_t>													23	24																							
<ptrdiff_t>													25	26																							
<ptrdiff_t>													27	28																							
<ptrdiff_t>													29	30																							
<ptrdiff_t>													31	32																							
<ptrdiff_t>													33	34																							
<ptrdiff_t>													35	36																							
<ptrdiff_t>													37	38																							
<ptrdiff_t>													39	40																							
<ptrdiff_t>													41	42																							
<ptrdiff_t>													43	44																							
<ptrdiff_t>													45	46																							
<ptrdiff_t>													47	48																							
<ptrdiff_t>													49	50																							
<ptrdiff_t>													51	52																							
<ptrdiff_t>													53	54																							
<ptrdiff_t>													55	56																							
<ptrdiff_t>													57	58																							
<ptrdiff_t>													59	60																							
<ptrdiff_t>													61	62																							

## Lexikálna analýza

---

Riešenie lexikálnej analýzy sme začali vytvorením diagramu deterministického konečného automatu. Následne sme na jeho základe začali vypracovávať implementáciu. Implementácia sa nachádza v súbore *scanner.c*, ktorý pracuje s tokenmi deklarovanými v súbore *token.h*. Hlavnou funkciou *scanner.c* je funkcia *get\_token*. Pre uľahčenie práce a prehľadnosti kódu sme si deklarovali niekoľko makier, ktoré sú extensívne používané v hlavnej funkcii. Funkcia *get\_token* berie postupne znaky zo štandardného vstupu a vytvára token. Tokenu je priradený jeho typ a hodnota, ktorá mu odpovedá. Funkcia začína určovaním jednoznakových tokenov, ktoré vie určiť hneď na začiatku. Pokračuje identifikáciou komentárov, ktoré následne ignoruje. Po identifikácii komentárov zisťuje či sa jedna o ID alebo Keyword, pri kľúčových slovách sa následne určuje aj ich typ. Ak sa nejedná ani o jedno pokračuje kontrolou dátových typov, pri ktorých ukladá aj ich hodnoty.

## Syntaktická analýza

---

Riešenie syntaktickej analýzy sme započali vytvorením LL gramatiky, LL tabuľky a precedenčnej tabuľky. Následne na ich základe sme vypracovali súbor *parser.c* a *exp\_parser.c*. Tieto súbory pracujú s uzlami deklarovanými v súbore *ast.h*. Spustenie syntaktickej analýzy započne zavolaním funkcie *Parse()*. Tato funkcia postupne prechádza cez tokeny a priradzuje ich do uzlov, pomocou ktorých postupne tvorí abstraktný syntaktický strom na základe LL gramatiky. Súbor *parser.c* ďalej riadi aj precedenčnú analýzu volaním funkcii zo súboru *exp\_parser.c*. Tento súbor vytvorí strom výrazov, ktorý je následne pripojený do syntaktického stromu.

## Semantická analýza

---

Sémantická analýza je implementovaná v súboroch *sem\_anal.c*, *symtable.c*, *sem\_anal.h* a *symtable.h*. Spustenie sémantickej analýzy započne zavolaním funkcie *analyse()*. Sémantická analýza je založená na rekurzívnom prechode AST stromu, ktorý je prevzatý od funkcie *parse()*, ktorá ho vygenerovala. Funkcia ďalej využíva globálne deklarovaný AST strom, v ktorom sa nachádzajú built in funkcie. Pri generácii nového AST stromu sa do neho vkladajú built in funkcie práve z tohto globálneho stromu. Funkcia *analyse()* po spustení hľadá *main* a následne postupne rekurzívne prechádza cez AST strom, kde kontroluje validitu dátových typov uložených v ASTNode štruktúrach. Po tejto kontrole započne aj kontrola navratových hodnôt. Po úspešnej validácii dát predáva nový AST strom funkcii *codegen()*, ktorá začína generáciu kódu. Ak validácia neprebehne



úspešne, vyhlási sematicku chybu.

Symtable, ktorý tato funkcia využíva je implementovaný ako AVL strom.

## Generovanie kodu

---

Generátor je implementovaný v súboroch *codegen\_priv.h*, *codegen.h* a *codegen.c*. Spustenie generácie kódu započne zavolaním funkcie *codegen()*. Kód je generovaný na základe rekurzívneho prechádzania abstraktného syntaktického stromu podľa dátového typu uloženého v štruktúre *ASTNode*. Kód ďalej využíva aj pomocný Linked List na ukladanie deklarovaných premenných v danej funkcii.

# Dátové štruktúry

---

## Character Buffer

Implementovane v súboroch *c\_buff.c* *c\_buff.h*.

Implementácia Character Buffer je využitá hlavne v časti Scanner, kde slúži na bezproblémové získavanie dát a ich následnú validáciu. Na prácu so scannerom ho neskôr využívajú aj časti Parser a Expression Parser. Štruktúra obsahuje klasické funkcie *c\_buff\_init*, *c\_buff\_free*, *c\_buff\_enqueue*, *c\_buff\_dequeue*, *c\_buff\_is\_empty*.

## Dynamic String

Implementovane v súboroch *dyn\_str.c*, *dyn\_str.h*.

Implementácia dynamického reťazca je využitá hlavne v Scanner časti programu, kde sprostredkúva validáciu a uschovávanie dát, neskôr je použitá aj v časti Codegen, kde slúži na uľahčenie validácie dát. Štruktúra dynamického reťazca obsahuje klasické funkcie *dyn\_str\_init*, *dyn\_str\_grow*, *dyn\_str\_append*, *dyn\_str\_append\_str* a *dyn\_str\_free*.

## Stack

Implementovane v súboroch *stack.c*, *stack.h*.

Implementáciu nášho zásobníku využívame v Expression Parser časti programu. Štruktúra zásobníku je implementovaná s klasickými funkciami *stackInit*, *stackPush*, *stackPop*, *stackIsEmpty*, *stackClear* a *stackGetTop*. Zásobník sme zvolili pre jeho optimálny prístup k dátam a zachovanie jednoduchosti kódu.

## Poznámky

---

### BOOLTHEN

Nepodporuje :

- Viac relačných operátorov za sebou