

Projektová dokumentácia

Implementácia prekladača imperatívneho jazyka IFJ21

Tým 082, varianta II

8. decembra 2021

Andrei Shchapaniak	(xshcha00)	25 %
Andrej Binovsky	(xbinov00)	25 %
Zdenek Lapes	(xlapes02)	25 %
Richard Gajdosik	(xgajdo33)	25 %

Obsah

1	Úvod	1
2	Návrh a implementácia	1
2.1	Lexikálna analýza	1
2.2	Syntaktická analýza	1
2.3	Spracovanie výrazov pomocou precedenčnej syntaktickej analýzy	1
2.3.1	Implementácia precedenčnej tabuľky	1
2.4	Sémantická analýza	2
2.5	Generovanie cieľového kódu	2
2.5.1	Implementácia výpisu cieľového kódu	2
2.5.2	Generovanie – Deklarácie premenných	2
2.5.3	Generovanie – funkcie	2
2.5.4	Generovanie – výrazy	2
2.5.5	Generovanie – podmienky a cykly	2
2.6	Prekladový systém	3
2.6.1	CMake	3
2.6.2	Skripty	3
2.6.3	GNU Make	3
3	Špeciálne algoritmy a dátové štruktúry	3
3.1	Tabuľka s rozptýlenými položkami	3
3.2	Pole rozptýlených tabuliek	3
3.3	Obojsmerný rad	3
3.4	Dynamický reťazec	3
4	Práca v tímu	4
4.0.1	Komunikácia a spôsob práce v tíme	4
4.0.2	Verzovací systém a vývojové prostredie	4
4.1	Rozdelenie práce medzi členmi tímu	4
5	Záver	5

1 Úvod

2 Návrh a implementácia

2.1 Lexikálna analýza

Scanner slúži pre lexikálnu analýzu. Je implementovaný ako deterministický konečný automat, ktorý rozpoznáva všetky prichádzajúce tokeny. Uchováva informácie o tom či sa jedná o komentár, identifikátor, textové bloky, relačné operátory alebo iné validné, poprípade nevalidné tokeny u ktorých nastane lexikálna chyba 1. V prípade validného tokenu na základe koncového stavu sa vyplnia nasledujúce informácie v štruktúre `token_t`:

- `type` – typ načítaného tokenu
- `keyword` – keď typ je `T_KW`, do premennej `keyword` sa uloží odpovedajúca hodnota
- `attr` – hodnota tokenu
- `attr.num_i` – integer number
- `attr.num_f` – double number
- `attr.id` – ostatné tokeny

V prípade že sa narazí na komentár je celý textový blok, ktorý je podľa správnej lexikálnej štruktúry chápaný ako komentár ignorovaný.

2.2 Syntaktická analýza

Parser je hlavným modulom prekladača, pretože komunikuje se všetkými ostatnými modulmi a riadi celú funkčnosť prekladača. Syntaktická analýza sa vykonáva zhora dolu metódou rekurzívneho zostupu. Syntaktická analýza dostáva postupne od lexikálneho analyzátoru tokeny, ktoré následne musia spĺňať presnú syntaktickú štruktúru a postupnosť podľa pravidiel LL-gramatiky. V prípade porušenia pravidla (typ prichádzajúceho tokenu se líši od očakávaného) sa vyhodí syntaktická chyba 2. V priebehu syntaktickej analýzy sú volané funkcie z modulu pre generovanie kódu, ktoré vygenerujú cieľový kód z jazyka `Teal` do `IFJcode21`, ktorý následne spracováva interpret.

2.3 Spracovanie výrazov pomocou precedenčnej syntaktickej analýzy

Precedenčná syntaktická analýza je modul ktorý zaisťuje spracovanie výrazov metódou zdola hore. Vo svojom rozhraní obsahuje `expression()`, ktorú volá parser, keď chce precedenčnej analýze predať riadenie vo chvíli, kedy očakáva výraz.

2.3.1 Implementácia precedenčnej tabuľky

Postupne precedenčná analýza spracováva tokeny a pomocou precedenčnej tabuľky symbolov určuje precedenciu. Na základe tejto precedencie môže nastať 5 stavov:

1. Pri precedencii `<` pridávame na zásobník načítaný token spolu so znakom precedencie.
2. Pri precedencii `>` redukuje dva výrazy na jeden a ukladáme ich typ podľa pravidiel.
3. Pri precedencii `=` zapíšeme načítaný znak z tokenu na zásobník.
4. Pri precedencii `c` sme narazili na nesprávne poradie znakov a nastáva sémantická chyba.
5. Pri precedencii `s` sme narazili na dva identifikátory, redukuje zvyšok výrazu a vraciame parseru riadenie a výsledný typ.

2.4 Sémantická analýza

- Sémantické chyby pre nekompatibilitu typu priradenia, návratových hodnôt a predaných argumentov do funkcií sa detekujú pomocou dvoch polí (ľavá strana typu a pravá strana typu). Po spracovaní určitého pravidla sa vykoná porovnanie podľa sémantiky jazyka IFJ21.
- Kompatibilita typu vo výraze sa detekuje tým istým spôsobom, ale je použitý dve premenné typu char.
- Sémantické chyby pre nedefiníciu, redefiníciu sa detekujú pomocou tabuliek symbolov. Globálna tabuľka symbolov je určená pre názvy funkcií. Pre názvy premenných je vytvorený zásobník tabuliek symbolov. Dôvodom implementácie zásobníka je riešenie problému s totožnými názvami premenných v rôznych rámcoch.

2.5 Generovanie cieľového kódu

Generovanie cieľového kódu `IFJcode21` je implementovaný ako samostatný modul, ktorý je riadený syntaxou. Komponenty modulu sú volané v parseri na základe pravidiel `LL-gramatiky`. Z dôvodu neoptimalizácie sa cieľový kód generuje priamo bez tvorby trojadresného kódu.

2.5.1 Implementácia výpisu cieľového kódu

Na zaistenie výpisu cieľového kódu len za podmienky bezchybnej analýzy zapisujeme cieľový kód do dvoch textových blokov – definície funkcií a volanie funkcií. Tieto dva textové bloky po úspešnej analýze skonkatenujeme a vypíšeme na štandardný výstup.

2.5.2 Generovanie – Deklarácie premenných

Deklarácie premenných a ich možný konflikt názvov (na základe výskytu toho istého názvu v rôznych rámcoch) sme implementovali vďaka obojsmernému radu v ktorom sa ukladá adresa elementu tabuľky symbolov s príslušným identifikátorom. Element tabuľky obsahuje unikátne číslo premennej, ktorý zaisťuje jedinečnosť názvu premennej.

2.5.3 Generovanie – funkcie

Volanie funkcií je zaistené vygenerovaním kódu, ktorý predá funkcii argumenty pomocou dočasného rámcu. Následne je vygenerovaný kód pre zavolanie funkcie. Pre správnu funkčnosť volanej funkcie je ihneď na začiatku generovaný kód, ktorý z dočasného rámcu vytvorí lokálny rámec funkcie a pre všetky argumenty ktoré boli funkcii predané vytvorí premenné s názvami podľa parametrov funkcie. Následne sa generuje kód tela funkcie.

2.5.4 Generovanie – výrazy

Generovanie kódu pre výraz sa začne vykonávať ihneď po jeho redukcii. V priebehu redukcii je výraz zapísaný do obojstranného radu v postfixovom formáte. Jednotlivé elementy v rade nesú všetky potrebné informácie na generovanie výrazu - typ operátora, názov premennej či hodnotu konštanty. Generátor generuje len inštrukcie kódu `IFJcode21` ktoré využívajú zásobník. To znamená že hodnoty, medzivýsledky a následne výsledok výrazu sú uložené na zásobník.

2.5.5 Generovanie – podmienky a cykly

Pre generovanie podmienok a cyklov využívame návštevia, ktoré sú taktiež reprezentované unikátnym číslom a názvom funkcie kde sa nachádzajú. Na zabránenie redeklarácie premenných sa telo cyklu zapisuje do dvoch rôznych textových blokov. Vyskytnuté deklarácie zapisujeme naďalej do bloku definícií funkcií. No zvyšný

kód tela cyklu zapisujeme do tretieho pomocného textového bloku. Následne po vygenerovaní celého cyklu tieto dva bloky skonkatenujeme.

2.6 Prekladový systém

2.6.1 CMake

CMake je multiplatformný nástroj na preklad zdrojových kódov. Nástroj sme vybrali na základe preferencií všetkých členov tímu. CMake nám predovšetkým pomáhal kompilovať a testovať výsledný program. Pravidlá pre preklad sú napísané v súbore `CMakeLists.txt` a po spustení nástroja CMake je automaticky vygenerovaný súbor `Makefile`. Na testovanie sme používali `google testy`, ktoré sme prekladali výhradne pomocou CMaku.

2.6.2 Skripty

Pre účely testovania boli vytvorené shellovské skripty. Jeden rozsiahly skript, ktorý uľahčoval a automatizoval testovanie všetkých častí projektu. Taktiež sme vytvorili skript pre preklad projektu a čistenie prebytočných súborov vzniknutých v dôsledku kompilácie projektu či testovacích suborov.

2.6.3 GNU Make

Zo zadania bolo požadované aby odovzdaný projekt obsahoval `Makefile`, ktorý s príkazom `make` preloží zdrojové súbory projektu a s príkazom `make clean` zmaže prebytočné súbory vzniknuté v dôsledku kompilácie. Tento nástroj nám taktiež pomáhal zabaliť celý projekt do jedného archívu `zip`.

3 Špeciálne algoritmy a dátové štruktúry

3.1 Tabuľka s rozptýlenými položkami

Túto dátovú štruktúru sme si zvolili vďaka časovej zložitosti (je reprezentovaná medzi $O(1)$ až $O(n)$). Taktiež nám pomohli znalosti štruktúry z predmetov `IAL` a `IJC`. Veľkosť tabuľky jsme zvolili 101. Ako unikátny kľúč pre prístup k dátam v tabuľke slúži názov identifikátoru a názvy funkcií. Každý záznam v tabuľke obsahuje informácie o identifikátore. U premenných je uchovávaná aj informácia o hĺbke (redefinícia v zanorenejšom rámci kódu). Modul je implementovaný v súboroch `syntable.h` a `syntable.c`.

3.2 Pole rozptýlených tabuliek

TODO Modul je implementovaný v súboroch `symstack.c` a `symstack.h`

3.3 Obojsmerný rad

Obojsmerný rad je kombinácia zásobníka a radu. Je možné do neho vkladať aj odoberať dáta z oboch strán. Implementovali sme ho ako samostatný modul pre viac častí projektu. Je využívaný najmä v generovaní kódu, kde slúži na uchovávanie postfixového výrazu či identifikátorov. Taktiež obsahuje informáciách o parametroch, argumentoch a návratových hodnotách funkcií. Modul je implementovaný v súboroch `queue.c` a `queue.h`.

3.4 Dynamický reťazec

Pre uchovanie vygenerovaného kódu počas prekladu a prácu s identifikátormi sme vytvorili štruktúru `string_t`. Pre obsluhu štruktúry sme vytvorili pomocné funkcie ako alokácia/dealokácia štruktúry, odstraňovanie, pridávanie a konkaténacia reťazcov. Modul je implementovaný v súboroch `str.c` a `str.h`.

4 Práca v tímu

Ihneď pri skladaní tímu sme si všetci uvedomovali, že sa očakáva pravidelná a skorá práca na projekte, čo sa nám nakoniec aj podarilo. Každý na projekte pracoval vždy s predstihom a darilo sa nam dodržiavať termíny, ktoré sme si stanovili.

4.0.1 Komunikácia a spôsob práce v tíme

Pre komunikáciu sme používali výhradne komunikačnú platformu `Discord`. Funguje na rovnakom princípe ako platforma `Slack`, ktorá sa používa profesionálne účely. Na danej platforme boli vytvorené komunikačné vlákna v ktorých boli založené `TODO` či `error` listy. Taktiež tam prebiehala bežná komunikácia či hlasové rozhovory s možnosťou zdieľania obrazovky, vďaka čomu sme mohli vyriešiť mnoho problémov digitálne a tým aj veľmi rýchlo. Avšak aj napriek dobrej digitálnej komunikácii sme sa snažili mať čo najviac osobných stretnutí.

4.0.2 Verzovací systém a vývojové prostredie

Ako vývojové prostredie sme využili `Clion` a `Vim`. Vývoj prebiehal na platformách `MacOS`, `Linux` a `Windows`. No testovanie prebiehalo len na operačnom systéme `Linux`. Ako verzovací systém sme použili `git` spolu s portálom `GitHub`.

4.1 Rozdelenie práce medzi členmi tímu

Andrei:

- Lexikálna, sémantická a obecná syntaktická analýza
- Organizácia a kontrola práce nad projektom
- Tabuľka symbolov

Richard:

- Syntaktická a sémantická analýza pre výrazy
- Precedenčná tabuľka
- Prezentácia

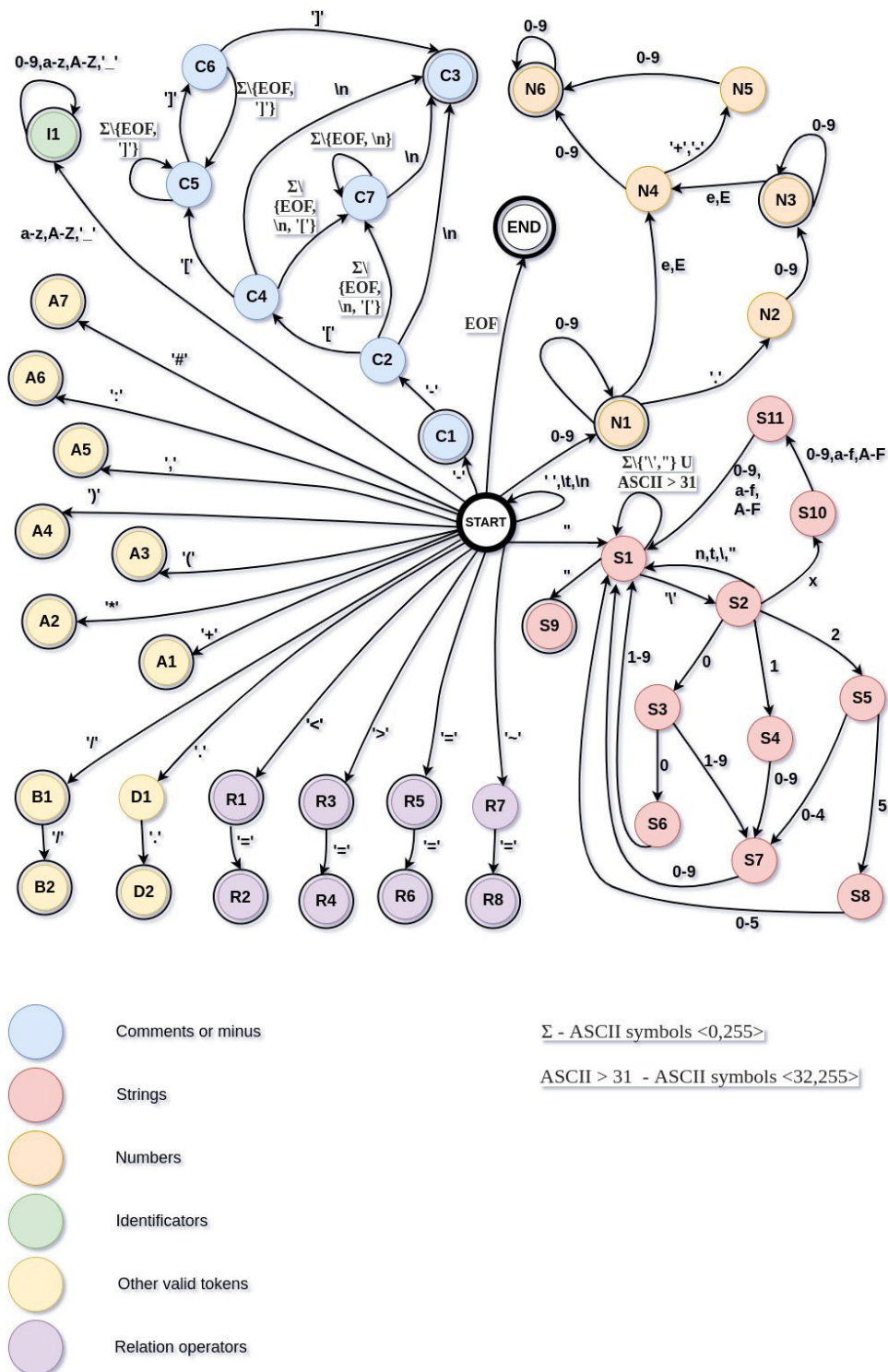
Zdenek a Andrej:

- Generovanie kódu
- Automatizácia testovania
- Google testy, tvorba testov
- Dokumentácia

5 Záver

Projekt bol zadaný začiatkom semestra, ešte skôr, než boli prebraté všetky potrebné znalosti pre úspešné splnenie projektu. Boli sme si ale vedomí zložitosti a obsiahlosti projektu, a preto sme čerpali informácie aj zo záznamov prednášok z minulých rokov. To nám dalo potrebné informácie do štartu projektu a tým sme mohli začať s implementáciou s dostatočne veľkým predstihom. Už začiatkom semestra sme mali zostavený tím. Taktiež sme sa dohodli na komunikačných kanáloch a rozdelení práce. Takže s komunikáciou neboli žiadne problémy a každý vedel, čo sa od neho očakáva. Ak sme mali nejaké nejasnosti so zadáním, tak všetko sme si vyjasnili buď pomocou diskusného fóra alebo neoficiálneho komunikačného kanála študentov. V implementácii projektu sme používali sadu testov (vlastných aj zdieľaných s kolegami). Tieto testy nam veľmi dobre poslúžili na detekovanie chýb a urýchlenie práce nad projektom. Projekt bol pre nás veľmi prínosná skúsenosť. Priamo v praxi sme si objasnili veľa problémov ohľadom prekladačov a naučili sme sa ako fungujú. Zároveň sme ako tím zvládli pracovať veľmi pekným tempom. Tímová pomoc bola samozrejmosťou.

Diagram konečného automatu specifikujúceho lexikálny analyzátor



Obr. 1: Diagram konečného automatu specifikujúci lexikálny analyzátor

LL – gramatika

1. $\langle \text{prolog} \rangle \rightarrow \text{require t_string } \langle \text{prog} \rangle$
2. $\langle \text{prog} \rangle \rightarrow \text{global id : function (} \langle \text{arg_T} \rangle \text{) } \langle \text{ret_T} \rangle \langle \text{prog} \rangle$
3. $\langle \text{prog} \rangle \rightarrow \text{function id (} \langle \text{arg} \rangle \text{) } \langle \text{ret_T} \rangle \langle \text{stmt} \rangle \text{ end } \langle \text{prog} \rangle$
4. $\langle \text{prog} \rangle \rightarrow \text{id (} \langle \text{param} \rangle \text{) } \langle \text{prog} \rangle$
5. $\langle \text{prog} \rangle \rightarrow \text{EOF}$
6. $\langle \text{arg_T} \rangle \rightarrow \langle \text{type} \rangle \langle \text{next_arg_T} \rangle$
7. $\langle \text{arg_T} \rangle \rightarrow \varepsilon$
8. $\langle \text{next_arg_T} \rangle \rightarrow , \langle \text{type} \rangle \langle \text{next_arg_T} \rangle$
9. $\langle \text{next_arg_T} \rangle \rightarrow \varepsilon$
10. $\langle \text{ret_T} \rangle \rightarrow : \langle \text{type} \rangle \langle \text{next_ret_T} \rangle$
11. $\langle \text{ret_T} \rangle \rightarrow \varepsilon$
12. $\langle \text{next_ret_T} \rangle \rightarrow , \langle \text{type} \rangle \langle \text{next_ret_T} \rangle$
13. $\langle \text{next_ret_T} \rangle \rightarrow \varepsilon$
14. $\langle \text{arg} \rangle \rightarrow \text{id : } \langle \text{type} \rangle \langle \text{next_arg} \rangle$
15. $\langle \text{arg} \rangle \rightarrow \varepsilon$
16. $\langle \text{next_arg} \rangle \rightarrow , \text{id : } \langle \text{type} \rangle \langle \text{next_arg} \rangle$
17. $\langle \text{next_arg} \rangle \rightarrow \varepsilon$
18. $\langle \text{type} \rangle \rightarrow \text{integer}$
19. $\langle \text{type} \rangle \rightarrow \text{number}$
20. $\langle \text{type} \rangle \rightarrow \text{string}$
21. $\langle \text{type} \rangle \rightarrow \text{nil}$
22. $\langle \text{stmt} \rangle \rightarrow \text{if } \langle \text{expr} \rangle \text{ then } \langle \text{stmt} \rangle \text{ else } \langle \text{stmt} \rangle \text{ end } \langle \text{stmt} \rangle$
23. $\langle \text{stmt} \rangle \rightarrow \text{while } \langle \text{expr} \rangle \text{ do } \langle \text{stmt} \rangle \text{ end } \langle \text{stmt} \rangle$
24. $\langle \text{stmt} \rangle \rightarrow \text{local id : } \langle \text{type} \rangle \langle \text{def_var} \rangle \langle \text{stmt} \rangle$
25. $\langle \text{stmt} \rangle \rightarrow \text{return } \langle \text{expr} \rangle \langle \text{next_expr} \rangle \langle \text{stmt} \rangle$
26. $\langle \text{stmt} \rangle \rightarrow \text{id } \langle \text{fork_id} \rangle \langle \text{stmt} \rangle$
27. $\langle \text{stmt} \rangle \rightarrow \varepsilon$
28. $\langle \text{def_var} \rangle \rightarrow = \langle \text{one_assign} \rangle$
29. $\langle \text{def_var} \rangle \rightarrow \varepsilon$
30. $\langle \text{one_assign} \rangle \rightarrow \text{id (} \langle \text{param} \rangle \text{) }$

31. $\langle \text{one_assign} \rangle \rightarrow \langle \text{expr} \rangle$
32. $\langle \text{param} \rangle \rightarrow \langle \text{param_val} \rangle \langle \text{next_param} \rangle$
33. $\langle \text{param} \rangle \rightarrow \varepsilon$
34. $\langle \text{param_val} \rangle \rightarrow \text{id}$
35. $\langle \text{param_val} \rangle \rightarrow \langle \text{term} \rangle$
36. $\langle \text{term} \rangle \rightarrow \text{t_string}$
37. $\langle \text{term} \rangle \rightarrow \text{t_integer}$
38. $\langle \text{term} \rangle \rightarrow \text{t_number}$
39. $\langle \text{term} \rangle \rightarrow \text{nil}$
40. $\langle \text{next_param} \rangle \rightarrow , \langle \text{param_val} \rangle \langle \text{next_param} \rangle$
41. $\langle \text{next_param} \rangle \rightarrow \varepsilon$
42. $\langle \text{next_expr} \rangle \rightarrow , \langle \text{expr} \rangle \langle \text{next_expr} \rangle$
43. $\langle \text{next_expr} \rangle \rightarrow \varepsilon$
44. $\langle \text{fork_id} \rangle \rightarrow (\langle \text{param} \rangle)$
45. $\langle \text{fork_id} \rangle \rightarrow \langle \text{next_id} \rangle$
46. $\langle \text{next_id} \rangle \rightarrow , \text{id} \langle \text{next_id} \rangle$
47. $\langle \text{next_id} \rangle \rightarrow = \langle \text{mult_assign} \rangle$
48. $\langle \text{mult_assign} \rangle \rightarrow \text{id} (\langle \text{param} \rangle)$
49. $\langle \text{mult_assign} \rangle \rightarrow \langle \text{expr} \rangle \langle \text{next_expr} \rangle$

LL – tabulka

	require	global	function	id	integer	string	number	nil	t_integer	t_number	t_string	if	while	local	return	:	=	(,	EOF	\$
<prolog>	1																				
<prog>		2	3	4																5	
<arg_T>					6	6	6	6													7
<next_arg_T>																			8		9
<ret_T>																10					11
<next_ret_T>																			12		13
<arg>				14																	15
<next_arg>																			16		17
<type>					18	20	19	21													
<stmt>				26								22	23	24	25						27
<def_var>																	28				29
<one_assign>				30																	31
<param>				32				32	32	32	32										33
<param_val>				34				35	35	35	35										
<term>								39	37	38	36										
<next_param>																			40		41
<next_expr>																			42		43
<fork_id>																	45	44	45		
<next_id>																	47		46		
<mult_assign>				48																	49

Precedenčná tabuľka

1. $E \rightarrow i$

2. $E \rightarrow (E)$

3. $E \rightarrow \# E$

4. $E \rightarrow E + E$

5. $E \rightarrow E - E$

6. $E \rightarrow E * E$

7. $E \rightarrow E / E$

8. $E \rightarrow E // E$

9. $E \rightarrow E .. E$

10. $E \rightarrow E > E$

11. $E \rightarrow E < E$

12. $E \rightarrow E >= E$

13. $E \rightarrow E <= E$

14. $E \rightarrow E == E$

15. $E \rightarrow E \sim = E$

	#	*	/	//	+	-	..	<	<=	>	>=	==	~=	()	i	\$
#	<	>	>	>	>	>	>	>	>	>	>	>	>	<	>	<	>
*	<	>	>	>	>	>	>	>	>	>	>	>	>	<	>	<	>
/	<	>	>	>	>	>	>	>	>	>	>	>	>	<	>	<	>
//	<	>	>	>	>	>	>	>	>	>	>	>	>	<	>	<	>
+	<	<	<	<	>	>	>	>	>	>	>	>	>	<	>	<	>
-	<	<	<	<	>	>	>	>	>	>	>	>	>	<	>	<	>
..	<	<	<	<	<	<	<	>	>	>	>	>	>	<	>	<	>
<	<	<	<	<	<	<	<	>	>	>	>	>	>	<	>	<	>
<=	<	<	<	<	<	<	<	>	>	>	>	>	>	<	>	<	>
>	<	<	<	<	<	<	<	>	>	>	>	>	>	<	>	<	>
>=	<	<	<	<	<	<	<	>	>	>	>	>	>	<	>	<	>
==	<	<	<	<	<	<	<	>	>	>	>	>	>	<	>	<	>
~=	<	<	<	<	<	<	<	>	>	>	>	>	>	<	>	<	>
(<	<	<	<	<	<	<	<	<	<	<	<	<	<	=	<	e
)	>	>	>	>	>	>	>	>	>	>	>	>	>	e	>	s	>
i	e	>	>	>	>	>	>	>	>	>	>	>	>	e	>	s	>
\$	<	<	<	<	<	<	<	<	<	<	<	<	<	<	e	<	e

LEGENDA:

< - insert to stack with shift

> - reduction

= - insert to stack

e - error

s - special case (end of expression)