

پروژه پایانی درس اصول طراحی کامپایلر
دانشگاه صنعتی جندی شاپور دزفول
دانشکده‌ی مهندسی برق و کامپیوتر - گروه مهندسی کامپیوتر
سال تحصیلی ۱۴۰۳-۱۴۰۲ - نیمسال دوم

عنوان پروژه: طراحی و پیاده‌سازی فازهای اول و دوم یک کامپایلر

شرح پروژه: هر گروه، بایستی با استفاده از یک زبان برنامه‌نویسی دلخواه و با بهره‌گیری از گرامر زبان برنامه‌سازی مورد نظر که در ادامه ارائه و تشریح می‌گردد، به پیاده‌سازی فازهای اول (تحلیل لغوی)^۱ و دوم (تحلیل نحوی)^۲ کامپایلر مورد نظر با زبان برنامه‌نویسی دلخواه بپردازد.

بخش اول - در این بخش، ابتدا لازم است کامپایلر مورد طراحی، متن کد نوشته شده در ویرایشگر^۳ را پویش^۴ نموده و نام، نوع و شماره سطر و ستون هر یک از توکن‌های مجاز را درون جدولی با نام جدول نمادها^۵ ذخیره نماید. در صورت غیرمجاز بودن هر توکن نیز پیغام مناسب به کاربر نمایش داده شود. حَسَب امکانات زبان برنامه‌نویسی مورد استفاده جهت توسعه کامپایلر، این جدول می‌تواند در قالب یک فایل متنی^۶ یا یک نمای بصری^۷ به عنوان اولین خروجی پروژه که مربوط به فاز تحلیل نحوی می‌باشد، ارائه شود.

بخش دوم - در این بخش، با استفاده از روش تحلیل نحوی بالا به پایین LL1 و با استفاده از گرامر داده شده، فرآیند بررسی تعلق کدهای نوشته شده به گرامر زبان بررسی می‌شود. به منظور افزایش کارایی فاز تحلیل نحوی، بهتر است توکن‌های مورد بررسی در این فاز، از جدول نمادها خوانده شود تا صرفاً توکن‌های مجاز مورد ارزیابی و تجزیه‌ی نحوی^۸ قرار گیرند. خروجی این فاز در صورت عدم وجود خطای نحوی، اجرای برنامه نوشته شده توسط کاربر خواهد بود؛ در صورت بروز هرگونه خطای نحوی، کامپایلر بایستی محل وقوع خطا را به کاربر نمایش دهد.

¹ Lexical Analysis

² Syntax Analysis

³ Editor

⁴ Scanning

⁵ Symbols Table

⁶ Text File

⁷ Visual View

⁸ Syntax Parsing

بخش اختیاری- به هر یک از گروه‌ها که بتواند با استفاده از روش تصحیح خطای حالت وحشت^۹، امکان تصحیح خطاهای نحوی رخ داده را فراهم نماید و نیز در هنگام بروز خطای نحوی، علاوه بر محل حادث شدن خطای نحوی، نوع آن را نیز گزارش نماید، نمره‌ی اضافی تعلق خواهد گرفت.

بخش فرضیات و داده‌های پروژه- در ادامه، گرامر مورد نظر جهت استفاده کامپایلر، فرم BNF، مجموعه‌های آغازین و پیرو و جدول تجزیه‌ی LL1 گرامر مورد نظر، ارائه شده است.

الف) گرامر زبان مورد طراحی:

<prog>	→	program <id> ; var <dec-list> begin <stat-list> end
<id>	→	<letter> { <letter> <digit> }
<dec-list>	→	<dec> : <type>
<dec>	→	<id> , <dec> <id>
<type>	→	integer
<stat-list>	→	<stat> <stat> <stat-list>
<stat>	→	<write> <assign>
<write>	→	show (<id>) ;
<assign>	→	<id> = <expr> ;
<expr>	→	<expr> + <term> <expr> - <term> <term>
<term>	→	<term> * <factor> <term> / <factor> <factor>
<factor>	→	<id> <number> (<expr>)
<number>	→	<sign> <digit> { <digit> }
<sign>	→	+ - λ
<digit>	→	0 1 2 3 4 5 6 7 8 9
<letter>	→	a b c d e

ب) فرم ^{۱۰}BNF گرامر^{۱۱}:

^۹ Panic Mode

^{۱۰} Backus-Naur Form

^{۱۱} برای اطلاعات بیشتر به تارنمای زیر رجوع شود:

https://fa.wikipedia.org/wiki/%D9%81%D8%B1%D9%85_%D8%A8%D8%A7%DA%A9%D9%88%D8%B3_%D9%86%D8%A7%D8%A6%D9%88%D8%B1_%D8%AA%D9%88%D8%B3%D8%B9%D9%87_%DB%8C%D8%A7%D9%81%D8%AA%D9%87

توجه شود که کلمات ضخیم شده (Bold) در فرم زیر، بیان گر پایانه‌های گرامر^{۱۲} می‌باشد.

(1)	<prog>	→	program <id> <;> var <dec-list> begin <stat-list> end
(2)	<id>	→	<letter> <alphaNum>
(3)	<alphaNum>	→	<letter> <alphaNum>
(4)	<alphaNum>	→	<digit> <alphaNum>
	<alphaNum>	→	λ
(5)	<dec-list>	→	<dec> : <type> <;>
(6)	<dec>	→	<id> <moreDec>
(7)	<moreDec>	→	<dec>
	<moreDec>	→	λ
(8)	<type>	→	integer
(9)	<stat-list>	→	<stat> <moreStat>
(10)	<moreStat>	→	<stat-list>
	<moreStat>	→	λ
(11)	<stat>	→	<write>
(12)	<stat>	→	<assign>
(13)	<write>	→	show <(> <id> <)> <;>
(14)	<assign>	→	<id> = <expr> <;>
(15)	<expr>	→	<term> <exprPrime>
(16)	<exprPrime>	→	+ <term> <exprPrime>
(17)	<exprPrime>	→	- <term> <exprPrime>
	<exprPrime>	→	λ
(18)	<term>	→	<factor> <termPrime>
(19)	<termPrime>	→	* <factor> <termPrime>
(20)	<termPrime>	→	/ <factor> <termPrime>
	<termPrime>	→	λ
(21)	<factor>	→	<id>
(22)	<factor>	→	<number>
(23)	<factor>	→	<(> <expr> <)>
(24)	<number>	→	<digit> <moreDigits>
(25)	<number>	→	<sign> <digit> <moreDigits>
(26)	<moreDigits>	→	<digit> <moreDigits>
	<moreDigits>	→	λ
(27)	<sign>	→	+

¹² Grammar Terminals

(28)	<sign>	→	-
(29)	<sign>	→	λ
(30)	<digit>	→	0
(31)	<digit>	→	1
(32)	<digit>	→	2
(33)	<digit>	→	3
(34)	<digit>	→	4
(35)	<digit>	→	5
(36)	<digit>	→	6
(37)	<digit>	→	7
(38)	<digit>	→	8
(39)	<digit>	→	9
(40)	<letter>	→	a
(41)	<letter>	→	b
(42)	<letter>	→	c
(43)	<letter>	→	d
(44)	<letter>	→	e
(45)	<;>	→	;
(46)	<(>	→	(
(47)	<)>	→)

پ) عناصر مجموعه‌های آغازین^{۱۳} و پیرو^{۱۴} گرامر:

Non-terminals	FIRST	FOLLOW
<prog>	program	\$
<id>	a...e	+ - * / =) , ; ;
<alphaNum>	a...e 0...9 λ	+ - * / =) , ; ;
<dec-list>	a...e	begin
<dec>	a...e	:
<moreDec>	λ	:
<type>	Integer	;
<stat-list>	show a...e	end
<stat>	show a...e	show end a...e
<moreStat>	show a...e λ	end
<write>	show	show end a...e

¹³ First Set

¹⁴ Follow Set

<assign>	a...e	show end a...e
<expr>	a...e 0...9 + - ();
<exprPrime>	+ - λ);
<term>	a...e + - (λ	+ -);
<termPrime>	* / λ	+ -);
<factor>	a...e + - (λ	+ - * /);
<number>	+ - λ	+ - * /);
<moreDigits>	0...9 λ	+ - * /);
<sign>	+ - λ	0...9
<digit>	0...9	0...9 + - * /);
<letter>	a...e	a...e 0...9
<;>	;	begin var show end a...e
<(>	(a...e 0...9 + - (
<)>)	+ - * /);

ت) جدول تجزیه LL1 گرامر:

جهت تسهیل در فرآیند پیاده‌سازی کامپایلر مورد نظر، جدول تجزیه LL1 گرامر مورد نظر در صفحه بعد ارائه شده است.

	program	var	begin	end	:	,	integer	show	=	+	-	*	/	0 – 9	a – e	;	()
<prog>	(1)																	
<id>															(2)			
<alphaNum>					λ	λ			λ	λ	λ	λ	λ	(4)	(3)	λ		λ
<dec-list>															(5)			
<dec>															(6)			
<moreDec>					λ	(7)												
<type>							(8)											
<stat-list>								(9)							(9)			
<moreStat>				λ				(10)							(10)			
<stat>								(11)							(12)			
<write>								(13)										
<assign>															(14)			
<expr>										(15)	(15)			(15)	(15)		(15)	
<exprPrime>										(16)	(17)					λ		λ
<term>										(18)	(18)			(18)	(18)		(18)	
<termPrime>										λ	λ	(19)	(20)			λ		λ
<factor>										(22)	(22)			(22)	(21)		(23)	
<number>										(25)	(25)							
<moreDigits>										λ	λ	λ	λ			λ		λ
<sign>										(27)	(28)							
<digit>														(30 – 39)				
<letter>															(40 – 44)			
<;>																(45)		
<(>																	(46)	
<>																		(47)

با آرزوی موفقیت و سربلندی- امین عنایت زارع