CS 4240 Phase 1

David Zhang

February 17, 2014

1 Lexical Rules

Pre-parse the input to remove comments with the following DFA. The resulting NOT-COMMENT tokens are concatenated together.

start state	symbol	next state	token
START	$\Sigma - \{/, ”\}$	START	NOT-COMMENT
START	“	STRING	
START	/	SLASH	
STRING	$\Sigma - \{\backslash, ”\}$	STRING	
STRING	\	STRING-SLASH	
STRING	”	START	NOT-COMMENT
STRING-SLASH	Σ	STRING	
SLASH	$\Sigma - \{*\}$	START	NOT-COMMENT
SLASH	*	COMMENT	
COMMENT	$\Sigma - \{*\}$	COMMENT	
COMMENT	*	COMMENT-END	
COMMENT-END	$\Sigma - \{*, /\}$	COMMENT	
COMMENT-END	*	COMMENT-END	
COMMENT-END	/	START	COMMENT

The DFA for uncommented code.

Note, any time backtracking is mentioned, it essentially is the same as not having read the last character and transitioning to the start state and then reading the last character. Also, the token produced does not include the last character. This is included to prevent duplication of the start state transitions.

Note, all ids are later matched character by character with keywords to determine if they are keywords.

Note, drop character for the error state means that the last read character is ignored, and the state remains unchanged.

Note, the accept state is also the start state. Failure to end in the accept state results in the current character buffer being abandoned, and an error message stating a incomplete state of the current state name.

start state	symbol	next state	token
START	+	START	PLUS
START	-	START	MIN
START	*	START	MULT
START	/	START	DIV
START	=	START	EQ
START	(	START	LPAREN
START	)	START	RPAREN
START	,	START	COMMA
START	&	START	AND
START	—	START	OR
START	[	START	LSQUARE
START	]	START	RSQUARE
START	;	START	SEMI
START	$\langle$	LANGLE	
START	$\rangle$	RANGLE	
START	:	COLON	
START	0-9	INT-LIT	
START	“	STRING-LIT	
START	a-zA-Z	ID	
START	whitespace	START	ignore
START	others	ERROR	drop character
LANGLE	$\rangle$	START	NOTEQ
LANGLE	=	START	LESSEQ
LANGLE	$\Sigma - \{=, \rangle\}$	START	LESS , backtrack
RANGLE	=	START	GREATEREQ
RANGLE	$\Sigma - \{=\}$	START	GREATER, backtrack
COLON	=	START	ASSIGN
COLON	$\Sigma - \{=\}$	START	COLON, backtrack
INT-LIT	0-9	INT-LIT	
INT-LIT	$\Sigma - 0 - 9$	START	INT-LIT, backtrack
ID	a-zA-Z0-9_	ID	
ID	$\Sigma - a - zA - Z0 - 9\_$	START	ID, backtrack
STRING-LIT	”	START	STRING-LIT
STRING-LIT	$\Sigma - \backslash$	STRING-LIT	
STRING-LIT	\	STRING-LIT-SLASH	
STRING-LIT-SLASH	n	STRING-LIT	
STRING-LIT-SLASH	t	STRING-LIT	
STRING-LIT-SLASH	”	STRING-LIT	
STRING-LIT-SLASH	\	STRING-LIT	
STRING-LIT-SLASH	^	STRING-LIT-CTL	
STRING-LIT-SLASH	0-9	STRING-LIT-CODE-1	
STRING-LIT-SLASH	whitespace	STRING-LIT-SPACE	ignore 2 characters
STRING-LIT-SLASH	others	ERROR	drop character
STRING-LIT-CTL	@A-Z[\]^_ .	STRING-LIT	
STRING-LIT-CTL	others	ERROR	drop character
STRING-LIT-CODE-1	0-9	STRING-LIT-CODE-2	
STRING-LIT-CODE-1	others	ERROR	drop character
STRING-LIT-CODE-2	0-9	STRING-LIT	
STRING-LIT-CODE-2	others	ERROR	drop character
STRING-LIT-SPACE	whitespace	STRING-LIT-SPACE	ignore
STRING-LIT-SPACE	\	STRING-LIT	ignore
STRING-LIT-SPACE	others	ERROR	drop character

2 Grammar Rules

Raw Grammar:

symbol	rule
⟨tiger-program⟩	let ⟨declaration-segment⟩ in ⟨stat-seq⟩ end
⟨declaration-segment⟩	⟨type-declaration-list⟩ ⟨var-declaration-list⟩ ⟨funct-declaration-list⟩
⟨type-declaration-list⟩	NULL
⟨type-declaration-list⟩	⟨type-declaration⟩ ⟨type-declaration-list⟩
⟨var-declaration-list⟩	NULL
⟨var-declaration-list⟩	⟨var-declaration⟩ ⟨var-declaration-list⟩
⟨funct-declaration-list⟩	⟨funct-declaration⟩ ⟨funct-declaration-list⟩
⟨type-declaration⟩	type id = ⟨type⟩ ;
⟨type⟩	⟨type-id⟩
⟨type⟩	array [INTLIT] of ⟨type-id⟩
⟨type-id⟩	int
⟨type-id⟩	string
⟨type-id⟩	id
⟨var-declaration⟩	var ⟨id-list⟩ : ⟨type-id⟩ ⟨optional-init⟩ ;
⟨id-list⟩	id
⟨id-list⟩	id , ⟨id-list⟩
⟨optional-init⟩	NULL
⟨optional-init⟩	:= ⟨const⟩
⟨funct-declaration⟩	function id (⟨param-list⟩) ⟨ret-type⟩ begin ⟨stat-seq⟩ end ;
⟨param-list⟩	NULL
⟨param-list⟩	⟨param⟩ ⟨param-list-tail⟩
⟨param-list-tail⟩	NULL
⟨param-list-tail⟩	, ⟨param⟩ ⟨param-list-tail⟩
⟨ret-type⟩	NULL
⟨ret-type⟩	: ⟨type-id⟩
⟨param⟩	id : ⟨type-id⟩
⟨stat-seq⟩	⟨stat⟩
⟨stat-seq⟩	⟨stat⟩ ⟨stat-seq⟩
⟨stat⟩	⟨lvalue⟩ := ⟨expr⟩ ;
⟨stat⟩	if ⟨expr⟩ then ⟨stat-seq⟩ endif ;
⟨stat⟩	if ⟨expr⟩ then ⟨stat-seq⟩ else ⟨stat-seq⟩ endif ;
⟨stat⟩	while ⟨expr⟩ do ⟨stat-seq⟩ enddo ;
⟨stat⟩	for id := ⟨expr⟩ to ⟨expr⟩ do ⟨stat-seq⟩ enddo ;
⟨stat⟩	id (⟨expr-list⟩) ;
⟨stat⟩	break ;
⟨stat⟩	return ⟨expr⟩ ;
⟨expr⟩	⟨const⟩
⟨expr⟩	⟨lvalue⟩
⟨expr⟩	- ⟨expr⟩
⟨expr⟩	⟨expr⟩ ⟨binary-operator⟩ ⟨expr⟩
⟨expr⟩	(⟨expr⟩)
⟨const⟩	INTLIT
⟨const⟩	STRLIT
⟨const⟩	nil
⟨expr-list⟩	NULL
⟨expr-list⟩	⟨expr⟩ ⟨expr-list-tail⟩
⟨expr-list-tail⟩	, ⟨expr⟩ ⟨expr-list-tail⟩
⟨expr-list-tail⟩	NULL
⟨lvalue⟩	id ⟨lvalue-tail⟩
⟨lvalue-tail⟩	[⟨expr⟩] ⟨lvalue-tail⟩
⟨lvalue-tail⟩	NULL