The University of Nottingham

SCHOOL OF COMPUTER SCIENCE

A LEVEL 3 MODULE, AUTUMN SEMESTER 2017–2018

COMPILERS

Time allowed TWO hours

Candidates may complete the front cover of their answer book and sign their desk card but must NOT write anything else until the start of the examination period is announced.

Answer ALL THREE questions

No calculators are permitted in this examination.

Dictionaries are not allowed with one exception. Those whose first language is not English may use a standard translation dictionary to translate between that language and English provided that neither language is the subject of this examination. Subject-specific translation directories are not permitted.

No electronic devices capable of storing and retrieving text, including electronic dictionaries, may be used.

DO NOT turn examination paper over until instructed to do so

ADDITIONAL MATERIAL: Appendix A, Appendix B

INFORMATION FOR INVIGILATORS: none

Question 1

(a) Explain and give examples of the following kinds of compile-time error:

- lexical error
- syntax error (context-free)
- contextual error

(6)

(b) Draw the *parse* (or *derivation*) tree for the following MiniTriangle fragment. The relevant grammar is given in Appendix A. Start from the production for "Command".

```
if x[i] < 100 then
    putint(k)
else
    i := (-i) - 1
```

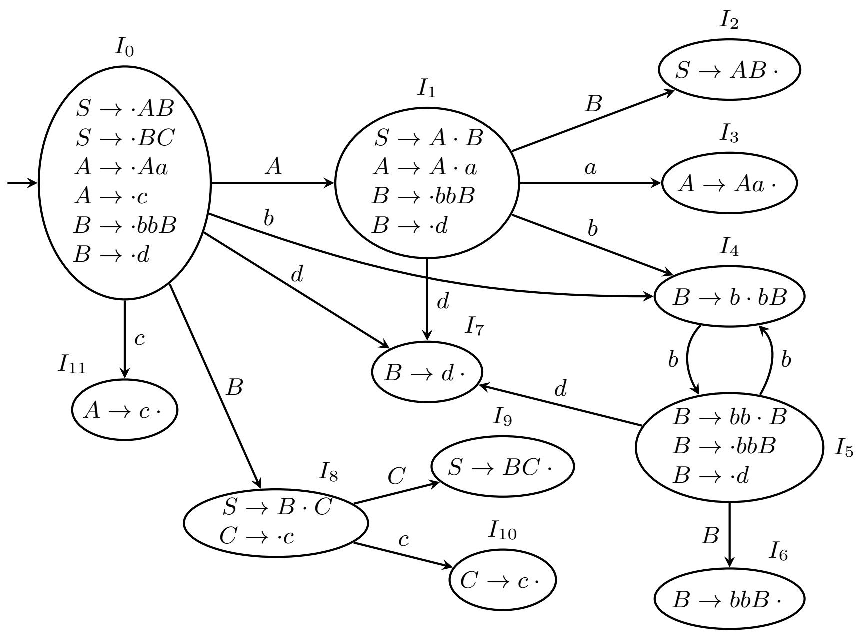
(9)

(c) Consider the following context-free grammar (CFG):

$$
\begin{array}{lcl}
S & \rightarrow & AB \mid BC \\
A & \rightarrow & Aa \mid c \\
B & \rightarrow & bbB \mid d \\
C & \rightarrow & c
\end{array}
$$

S, A, B, and C are nonterminal symbols, S is the start symbol, and a, b, c, and d are terminal symbols.

The DFA below recognizes the viable prefixes for this CFG:



Show how an LR(0) shift-reduce parser parses the string $caabbbbd$ by completing the following table (copy it to your answer book; do *not* write on the examination paper):

State	Stack	Input	Move
I_0	ϵ	$caabbbbd$	Shift
I_{11}	c	$aabbbbd$	Reduce by $A \rightarrow c$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
	S	ϵ	Done

(10)

Question 2
Consider the language given by the following abstract syntax:

$$
\begin{array}{lcll}
C & \rightarrow & & \textit{Commands:} \\
 & & \texttt{skip} & \textit{Do nothing} \\
 & | & C \texttt{ ; } C & \textit{Sequencing} \\
 & | & x \texttt{ := } E & \textit{Assignment} \\
 & | & \texttt{if } E \texttt{ then } C \texttt{ else } C & \textit{Conditional command} \\
 & | & \texttt{while } E \texttt{ do } C & \texttt{while}\textit{-loop} \\
\\
E & \rightarrow & & \textit{Expressions:} \\
 & & n & \textit{Literal integer} \\
 & | & x & \textit{Variable} \\
 & | & E \texttt{ + } E & \textit{Addition} \\
 & | & E \texttt{ = } E & \textit{Comparison}
\end{array}
$$

For this question, you will develop code generation functions for the above language, targeting the Triangle Abstract Machine (TAM). See appendix B for a specification of the TAM instructions. Assume conventional (imperative) semantics for the above language constructs, along with the following:

- x is the syntactic category of variable identifiers, ranging over the 26 names a, b, ..., z. They refer to 26 global variables stored at SB $+ 0$ (a) to SB $+ 25$ (z).
- The while-loop has the following semantics: the loop expression E is evaluated; if the result is true, the loop body C is executed next and then the process is repeated from the evaluation of the loop expression; otherwise execution continues after the loop.

The code generation functions should be specified through *code templates* in the style used in the lectures. Assume a function $addr(x)$ that returns the address (of the form [SB + d]) for a variable x. Further, you will have to consider generation of fresh labels. Assume a monadic-style operation $l \leftarrow \mathit{fresh}$ to bind a variable l to a distinct label that then can be used in jumps and as jump targets. For example:

$$
\begin{array}{rcrl}
\mathit{execute}\, [\![\, \texttt{if } E \texttt{ then } C_1 \texttt{ else } C_2 \,]\!] & = & & l_1 \leftarrow \mathit{fresh} \\
 & & & \ldots \\
 & & & \texttt{JUMP } l_1 \\
 & & & \ldots \\
 & & l_1: & \ldots
\end{array}
$$

(a) Write a code generation function $evaluate$ that generates TAM code for evaluating an expression. The first case should start like:

$$\mathit{evaluate}\, [\![\, n \,]\!] \quad = \quad \ldots$$

(4)

(b) Write a code generation function $execute$ that generates TAM code for executing commands. It should handle the five forms of commands specified by the abstract syntax above. (12)

(c) Now assume we wish to extend the language with the commands `break` and `continue` :

$$
\begin{array}{lll}
C & \rightarrow & & \textit{Commands:} \\
 & | & \ldots & \\
 & | & \texttt{break} & \textit{Terminate innermost loop} \\
 & | & \texttt{continue} & \textit{Continue with next loop iteration}
\end{array}
$$

The semantics is that `break` will terminate the innermost loop, with execution continuing immediately after the loop, while `continue` will skip whatever remains of the loop body, and continue execution directly with the next loop iteration.

Modify and extend $execute$ to generate code for the extended language. Note that $execute$ will need (an) extra argument(s) for contextual information to keep track of the current innermost loop. You may assume that using `break` or `continue` outside any loop is a *static* error. Thus your code generator does not need to handle that case. Your answer should include the modified $execute$ cases for `if` and `while`, as well as the cases for the two new commands. (9)

Question 3

This question concerns code improvement (optimisation) and internal representations that facilitate analysis and code improvement.

(a) Explain the code improvement technique *common subexpression elimination*, illustrating with an example. Also discuss when the technique cannot be applied, again illustrating with an example. (6)

(b) Show how the following program fragment involving a C-like `for`-loop might be transformed by means of *loop unrolling* in a situation where the loop bound `n` is *not* statically known:

```
b[0] := a[0];
for (i := 1; i < n; i++) do
    b[i] := b[i-1] + a[i];
```

Also discuss the potential advantages and disadvantages of this transformation. (9)

(c) Transform the following code fragment into *static single assignment* (SSA) form:

```
a := 0;
b := 1;
i := 2;
while i < n do begin
    c := a + b;
    a := b;
    b := c;
    i := i + 1
end
```

(10)

Appendix A: MiniTriangle Grammars

This appendix contains the grammars for the MiniTriangle lexical, concrete, and abstract syntax. The following typographical conventions are used to distinguish between terminals and non-terminals:

- nonterminals are written like *this*
- terminals are written like `this`
- terminals with *variable spelling* and special symbols are written like $\underline{this}$

MiniTriangle Lexical Syntax:

Program → (*Token* | *Separator*)*

Token → *Keyword* | *Identifier* | *IntegerLiteral* | *Operator*
| `,` | `;` | `:` | `:=` | `=` | `(` | `)` | `[` | `]` | $\underline{eot}$

Keyword → `begin` | `const` | `do` | `else` | `end` | `fun` | `if` | `in`
| `let` | `out` | `proc` | `then` | `var` | `while`

Identifier → *Letter* | *Identifier Letter* | *Identifier Digit*
except *Keyword*

IntegerLiteral → *Digit* | *IntegerLiteral Digit*

Operator → `^` | `*` | `/` | `+` | `-` | `<` | `<=` | `==` | `!=` | `>=` | `>` | `&&` | `||` | `!`

Letter → `A` | `B` | ... | `Z` | `a` | `b` | ... | `z`

Digit → `0` | `1` | `2` | `3` | `4` | `5` | `6` | `7` | `8` | `9`

Separator → *Comment* | $\underline{space}$ | $\underline{eol}$

Comment → `//` (any character except $\underline{eol}$)* $\underline{eol}$

MiniTriangle Concrete Syntax:

$$
\begin{array}{lcl}
\mathit{Program} & \rightarrow & \mathit{Command} \\
 & & \\
\mathit{Commands} & \rightarrow & \mathit{Command} \\
 & \mid & \mathit{Command}\ \texttt{;}\ \mathit{Commands} \\
 & & \\
\mathit{Command} & \rightarrow & \mathit{VarExpression}\ \texttt{:=}\ \mathit{Expression} \\
 & \mid & \mathit{VarExpression}\ \texttt{(}\ \mathit{Expressions}\ \texttt{)} \\
 & \mid & \texttt{if}\ \mathit{Expression}\ \texttt{then}\ \mathit{Command} \\
 & & \texttt{else}\ \mathit{Command} \\
 & \mid & \texttt{while}\ \mathit{Expression}\ \texttt{do}\ \mathit{Command} \\
 & \mid & \texttt{let}\ \mathit{Declarations}\ \texttt{in}\ \mathit{Command} \\
 & \mid & \texttt{begin}\ \mathit{Commands}\ \texttt{end} \\
 & & \\
\mathit{Expressions} & \rightarrow & \epsilon \\
 & \mid & \mathit{Expressions_1} \\
 & & \\
\mathit{Expressions_1} & \rightarrow & \mathit{Expression} \\
 & \mid & \mathit{Expression}\ \texttt{,}\ \mathit{Expressions_1} \\
 & & \\
\mathit{Expression} & \rightarrow & \mathit{PrimaryExpression} \\
 & \mid & \mathit{Expression}\ \mathit{BinaryOperator}\ \mathit{Expression} \\
 & & \\
\mathit{PrimaryExpression} & \rightarrow & \underline{\mathit{IntegerLiteral}} \\
 & \mid & \mathit{VarExpression} \\
 & \mid & \mathit{UnaryOperator}\ \mathit{PrimaryExpression} \\
 & \mid & \mathit{VarExpression}\ \texttt{(}\ \mathit{Expressions}\ \texttt{)} \\
 & \mid & \texttt{[}\ \mathit{Expressions}\ \texttt{]} \\
 & \mid & \texttt{(}\ \mathit{Expression}\ \texttt{)} \\
 & & \\
\mathit{VarExpression} & \rightarrow & \underline{\mathit{Identifier}} \\
 & \mid & \mathit{VarExpression}\ \texttt{[}\ \mathit{Expression}\ \texttt{]} \\
 & & \\
\mathit{BinaryOperator} & \rightarrow & \texttt{\^{}} \mid \texttt{*} \mid \texttt{/} \mid \texttt{+} \mid \texttt{-} \mid \texttt{<} \mid \texttt{<=} \mid \texttt{==} \mid \texttt{!=} \mid \texttt{>=} \mid \texttt{>} \mid \texttt{\&\&} \mid \texttt{||} \\
 & & \\
\mathit{UnaryOperator} & \rightarrow & \texttt{-} \mid \texttt{!}
\end{array}
$$

$$
\begin{array}{lcl}
\mathit{Declarations} & \rightarrow & \mathit{Declaration} \\
 & | & \mathit{Declaration}\ \texttt{;}\ \mathit{Declarations} \\
\\
\mathit{Declaration} & \rightarrow & \texttt{const}\ \underline{\mathit{Identifier}}\ \texttt{:}\ \mathit{TypeDenoter}\ \texttt{=}\ \mathit{Expression} \\
 & | & \texttt{var}\ \underline{\mathit{Identifier}}\ \texttt{:}\ \mathit{TypeDenoter} \\
 & | & \texttt{var}\ \underline{\mathit{Identifier}}\ \texttt{:}\ \mathit{TypeDenoter}\ \texttt{:=}\ \mathit{Expression} \\
 & | & \texttt{fun}\ \underline{\mathit{Identifier}}\ \texttt{(}\ \mathit{ArgDecls}\ \texttt{)}\ \texttt{:}\ \mathit{TypeDenoter}\ \texttt{=}\ \mathit{Expression} \\
 & | & \texttt{proc}\ \underline{\mathit{Identifier}}\ \texttt{(}\ \mathit{ArgDecls}\ \texttt{)}\ \mathit{Command} \\
\\
\mathit{ArgDecls} & \rightarrow & \epsilon \\
 & | & \mathit{ArgDecls}_1 \\
\\
\mathit{ArgDecls}_1 & \rightarrow & \mathit{ArgDecl} \\
 & | & \mathit{ArgDecl}\ \texttt{,}\ \mathit{ArgDecls}_1 \\
\\
\mathit{ArgDecl} & \rightarrow & \underline{\mathit{Identifier}}\ \texttt{:}\ \mathit{TypeDenoter} \\
 & | & \texttt{in}\ \underline{\mathit{Identifier}}\ \texttt{:}\ \mathit{TypeDenoter} \\
 & | & \texttt{out}\ \underline{\mathit{Identifier}}\ \texttt{:}\ \mathit{TypeDenoter} \\
 & | & \texttt{var}\ \underline{\mathit{Identifier}}\ \texttt{:}\ \mathit{TypeDenoter} \\
\\
\mathit{TypeDenoter} & \rightarrow & \underline{\mathit{Identifier}} \\
 & | & \mathit{TypeDenoter}\ \texttt{[}\ \underline{\mathit{IntegerLiteral}}\ \texttt{]}
\end{array}
$$

Note that the productions for *Expression* make the grammar as stated above ambiguous. Operator precedence and associativity for the *binary* operators as defined in the following table are used to disambiguate:

Operator	**Precedence**	**Associativity**
`^`	1	right
`* /`	2	left
`+ -`	3	left
`< <= == != >= >`	4	non
`&&`	5	left
`||`	6	left

A precedence level of 1 means the highest precedence, 2 means second highest, and so on.

MiniTriangle Abstract Syntax: $\underline{Name} = \underline{Identifier} \cup \underline{Operator}$.

Program	$\rightarrow$	*Command*	Program
Command	$\rightarrow$	*Expression* `:=` *Expression*	CmdAssign
	$\mid$	*Expression* `(` *Expression** `)`	CmdCall
	$\mid$	`begin` *Command** `end`	CmdSeq
	$\mid$	`if` *Expression* `then` *Command* `else` *Command*	CmdIf
	$\mid$	`while` *Expression* `do` *Command*	CmdWhile
	$\mid$	`let` *Declaration** `in` *Command*	CmdLet
Expression	$\rightarrow$	$\underline{IntegerLiteral}$	ExpLitInt
	$\mid$	$\underline{Name}$	ExpVar
	$\mid$	*Expression* `(` *Expression** `)`	ExpApp
	$\mid$	`[` *Expression** `]`	ExpAry
	$\mid$	*Expression* `[` *Expression* `]`	ExpIx
Declaration	$\rightarrow$	`const` $\underline{Name}$ `:` *TypeDenoter* `=` *Expression*	DeclConst
	$\mid$	`var` $\underline{Name}$ `:` *TypeDenoter* `(` `:=` *Expression* $\mid \epsilon$ `)`	DeclVar
	$\mid$	`fun` $\underline{Name}$ `(` *ArgDecl** `)` `:` *TypeDenoter* `=` *Expression*	DeclFun
	$\mid$	`proc` $\underline{Name}$ `(` *ArgDecl** `)` *Command*	DeclProc
ArgDecl	$\rightarrow$	*ArgMode* $\underline{Name}$ `:` *TypeDenoter*	ArgDecl
ArgMode	$\rightarrow$	ϵ	ByValue
	$\mid$	`in`	ByRefIn
	$\mid$	`out`	ByRefOut
	$\mid$	`var`	ByRefVar
TypeDenoter	$\rightarrow$	$\underline{Name}$	TDBaseType
	$\rightarrow$	*TypeDenoter* `[` $\underline{IntegerLiteral}$ `]`	TDArray

Appendix B: Triangle Abstract Machine (TAM) Instructions

Meta variable	Meaning
a	Address: one of the forms specified by table below when part of an instruction, specific stack address when on the stack
b	Boolean value ($\mathrm{false} = 0$ or $\mathrm{true} = 1$)
ca	Code address; address to routine in the code segment
d	Displacement; i.e., offset w.r.t. address in register or on the stack
l	Label name
m, n, p	Integer
x, y, z	Any kind of stack data
x^n	Vector of n items, $n \geq 0$, here any kind

Address form	Description
`[SB + `d`]` `[SB - `d`]`	Address given by contents of register `SB` (Stack Base) $+/-$ displacement d
`[LB + `d`]` `[LB - `d`]`	Address given by contents of register `LB` (Local Base) $+/-$ displacement d
`[ST + `d`]` `[ST - `d`]`	Address given by contents of register `ST` (Stack Top) $+/-$ displacement d

Instruction	Stack effect	Description
	Label	
`LABEL` l	—	Pseudo instruction: symbolic location
	Load and store	
`LOADL` n	$\ldots \Rightarrow n, \ldots$	Push literal integer n onto stack
`LOADCA` l	$\ldots \Rightarrow \mathrm{addr}(l), \ldots$	Push address of label l (code segment) onto stack
`LOAD` a	$\ldots \Rightarrow [a], \ldots$	Push contents at address a onto stack
`LOADA` a	$\ldots \Rightarrow a, \ldots$	Push address a onto stack
`LOADI` d	$a, \ldots \Rightarrow [a+d], \ldots$	Load indirectly; push contents at address $a + d$ onto stack
`STORE` a	$n, \ldots \Rightarrow \ldots$	Pop value n from stack and store at address a
`STOREI` d	$a, n, \ldots \Rightarrow \ldots$	Store indirectly; store n at address $a+d$

Instruction	Stack effect	Description
	Block operations	
LOADLB m n	$\ldots \Rightarrow m^n, \ldots$	Push block of n literal integers m onto stack
LOADIB n	$a, \ldots \Rightarrow [a+(n-1)], \ldots, [a+0], \ldots$	Load block of size n indirectly
STOREIB n	$a, x^n, \ldots \Rightarrow \ldots$	Store block of size n indirectly
POP m n	$x^m, y^n, \ldots \Rightarrow x^m, \ldots$	Pop n values below top m values
	Arithmetic operations	
ADD	$n_2, n_1, \ldots \Rightarrow n_1+n_2, \ldots$	Add n_1 and n_2, replacing n_1 and n_2 with the sum
SUB	$n_2, n_1, \ldots \Rightarrow n_1-n_2, \ldots$	Subtract n_2 from n_1, replacing n_1 and n_2 with the difference
MUL	$n_2, n_1, \ldots \Rightarrow n_1 \cdot n_2, \ldots$	Multiply n_1 by n_2, replacing n_1 and n_2 with the product
DIV	$n_2, n_1, \ldots \Rightarrow n_1/n_2, \ldots$	Divide n_1 by n_2, replacing n_1 and n_2 with the (integer) quotient
NEG	$n, \ldots \Rightarrow -n, \ldots$	Negate n, replacing n with the result
	Comparison & logical operations (false $=0$, true $=1$)	
LSS	$n_2, n_1, \ldots \Rightarrow n_1 < n_2, \ldots$	Check if n_1 is smaller than n_2, replacing n_1 and n_2 with the Boolean result
EQL	$n_2, n_1, \ldots \Rightarrow n_1 = n_2, \ldots$	Check if n_1 is equal to n_2, replacing n_1 and n_2 with the Boolean result
GTR	$n_2, n_1, \ldots \Rightarrow n_1 > n_2, \ldots$	Check if n_1 is greater than n_2, replacing n_1 and n_2 with the Boolean result
AND	$b_2, b_1, \ldots \Rightarrow b_1 \wedge b_2, \ldots$	Logical conjunction of b_1 and b_2, replacing b_1 and b_2 with the Boolean result
OR	$b_2, b_1, \ldots \Rightarrow b_1 \vee b_2, \ldots$	Logical disjunction of b_1 and b_2, replacing b_1 and b_2 with the Boolean result
NOT	$b, \ldots \Rightarrow \neg b, \ldots$	Logical negation of b, replacing b with the result

Instruction	Stack effect	Description
Control transfer		
JUMP l	—	Jump unconditionally to location identified by label l
JUMPIFZ l	$n, \ldots \Rightarrow \ldots$	Jump to location identified by label l if $n = 0$ (i.e., n is false)
JUMPIFNZ l	$n, \ldots \Rightarrow \ldots$	Jump to location identified by label l if $n \neq 0$ (i.e., n is true)
CALL l	$\ldots \Rightarrow$ PC $+ 1$, LB, 0, $\ldots$	Call global subroutine at location l: Activation record set up by pushing static link (0 for global level), dynamic link (value of LB), and return address (PC$+1$, address of instruction after the call instruction) onto the stack; PC $= l$; LB $=$ start of activation record (address of static link)
CALLI	$ca, sl, \ldots \Rightarrow$ PC $+ 1$, LB, $sl, \ldots$	Call subroutine indirectly: address of routine (ca) and static link to use (sl) on top of the stack; activation record and new PC and LB as for CALL
RETURN m n	$x^m, y^p, ra, olb, sl, y^n, \ldots \Rightarrow x^m, \ldots$	Return from subroutine, replacing activation record by result, jumping to return address (PC $= ra$), and restoring the old local base (LB $= olb$)
Input/Output		
PUTINT	$n, \ldots \Rightarrow \ldots$	Print n to the terminal as a decimal integer
PUTCHR	$n, \ldots \Rightarrow \ldots$	Print the character with character code n to the terminal
GETINT	$\ldots \Rightarrow n, \ldots$	Read decimal integer n from the terminal and push onto the stack
GETCHR	$\ldots \Rightarrow n, \ldots$	Read character from the terminal and push its character code n onto the stack
TAM Control		
HALT	—	Stop execution and halt the machine