Abgabetermin: Mi in der KW 21

☐ **Gr. 1,** Dr. G. Kronberger **Name** _____ **Aufwand in h** _____
☐ **Gr. 2,** Dr. H. Gruber
☐ **Gr. 3,** Dr. D. Auer **Punkte** _____ **Kurzzeichen Tutor / Übungsleiter** _____ / _____

1. MidiPascal (10 Punkte)

MiniPascal ist eine ziemlich "schwache" Sprache, da man mit ihr nicht "alle" Probleme lösen kann – sofern es überhaupt (Programmier-)Sprachen gibt, mit denen man alle ... ;-) Wesentliche Sprachkonstrukte, die MiniPascal fehlen, sind Verzweigungen und Schleifen. Also erweitern wir MiniPascal um die binäre Verzweigung (*IF*-Anweisung), die Abweisschleife (*WHILE*-Schleife) sowie die Verbundanweisung (*BEGIN ... END*) – und taufen die neue Sprache MidiPascal.

Nachdem wir mit dem Datentyp *INTEGER* und ohne Erweiterungen der Ausdrücke um relationale Operatoren auskommen wollen, verwenden wir für Bedingungen in Verzeigungen und Schleifen *INTEGER*-Variablen mit der Semantik, dass jeder Wert ungleich 0 als *TRUE* und (nur) der Wert 0 als *FALSE* interpretiert wird – so wie das z. B. in der Programmiersprache C definiert ist. Folgende Tabelle zeigt zur Verdeutlichung eine Abbildung von MidiPascal auf (vollständiges) Pascal:

MidiPascal	(vollständiges) Pascal
`VAR x: INTEGER;`	`VAR x: INTEGER;`
`IF x THEN ...`	**`IF x <> 0 THEN ...`**
`WHILE x DO ...`	**`WHILE x <> 0 DO ...`**

Mit diesen Spracherweiterungen könnte man dann z. B. ein MidiPascal-Programm schreiben, das für eine vom Benutzer / von der Benutzerin eingegebene Zahl *n* die Fakultät *f* = *n*! berechnet und diese ausgibt. Siehe Quelltextstück rechts.

```
f := n; n := n - 1;
WHILE n DO BEGIN
  f := n * f;
  n := n - 1;
END;
WRITE(f);
```

Damit diese neuen Sprachkonstrukte im Compiler umgesetzt werden können, sind zwei neue Bytecode-Befehle notwendig. Folgende Tabelle erläutert diese beiden Befehle:

Bytecode-Befehl	Semantik
`Jmp addr`	Springe an die Codeadresse *addr*
`JmpZ addr`	Hole oberstes Element vom Stapel, wenn dieses 0 (*zero*) ist, springe nach *addr*

Nun muss man nur noch klären, welche Bytecodestücke für die einzelnen, neuen MidiPascal-Anweisungen zu erzeugen sind. Folgende Tabelle stellt die notwendigen Transformationen anhand von Mustern dar:

MidiPascal	Bytecode (mit fiktiven Adressen)

`IF x THEN BEGIN`
`  then stats`
`END;`
`...`
	`  1  LoadVal x`
	`  4  JmpZ 99`
	`...  code for then stats`
	
	`99  ...`

MidiPascal	Bytecode (mit fiktiven Adressen)
	1 LoadVal x
IF x THEN BEGIN	4 **JmpZ 66**
then stats	... *code for then stats*
END	... **Jmp 99**
ELSE BEGIN	
else stats	**66** *code for else stats*
END;	
...	**99** ...
	1 LoadVal x
WHILE x DO BEGIN	4 **JmpZ 99**
while stats	... *code for while stats*
END	... **Jmp 1**
	99 ...

Bei der Implementierung dieser neuen Sprachkonstrukte tritt das Problem auf, für die Bedingungen auch Sprunganweisungen „nach unten“ erzeugen zu müssen, wobei die Zieladressen der Sprünge noch nicht bekannt sind. Dieses Problem kann mit dem so genannten *Anderthalbpass-Verfahren* gelöst werden: Man erzeugt zuerst eine Sprunganweisung mit einer fiktiven Adresse (z. B. 0) und korrigiert diese später, sobald die Zieladresse bekannt ist (mittels *FixUp*).

Im Moodle-Kurs finden Sie *ForMidiPascalCompiler.zip* einen um die beiden neuen Bytecode-Befehle und zwei neue Operationen (*CurAddr* und *FixUp*) erweiterten Code-Generator (*CodeDef.pas* und *CodeGen.pas*) und eine erweiterte MidiPascal-Maschine (*CodeInt.pas*), welche die neuen Befehle ausführen kann. Sie müssen nur mehr den lexikalischen Analysator um die neuen Schlüsselwörter und den Syntaxanalysator mit seinen semantischen Aktionen um die neuen Anweisungen erweitern. Verwenden Sie als Basis dazu folgenden Ausschnitt der ATG für MidiPascal:

```
Stat = [ ... (*assignment, read, and write statement here, new ones below*)

  | 'BEGIN' StatSeq 'END'

  | 'IF' ident↑identStr          SEM IF NOT IsDecl(idenStr) THEN BEGIN
                                       SemError('variable not declared');
                                   END; (*IF*)
                                   Emit2(LoadValOpc, AddrOf(identStr));
                                   Emit2(JmpZOpc, 0); (*0 as dummy address*)
                                   addr := CurAddr - 2; ENDSEM
    'THEN' Stat
    [ 'ELSE'                     SEM Emit2(JmpOpc, 0); (*0 as dummy address*)
                                   FixUp(addr, CurAddr);
                                   addr := CurAddr - 2; ENDSEM
      Stat
    ]                            SEM FixUp(addr, CurAddr); ENDSEM

  | 'WHILE' ident↑identStr       SEM IF NOT IsDecl(identStr) THEN BEGIN
                                       SemError('variable not declared');
                                   END; (*IF*)
                                   addr1 := CurAddr;
                                   Emit2(LoadValOpc, AddrOf(identStr));
                                   Emit2(JmpZOpc, 0); (*0 as dummy address*)
                                   addr2 := CurAddr - 2; ENDSEM
    'DO' Stat                    SEM Emit2(JmpOpc, addr1);
                                   FixUp(addr2, CurAddr); ENDSEM

].
```

2. Optimierender MidiPascal-Compiler (2 + 4 + 4 + 4 Punkte)

Arithmetische Ausdrücke kann man wie folgt durch Binärbäume darstellen: aus dem Operator wird der Wurzelknoten, aus dem linken Operanden der linke und aus dem rechten Operanden der rechte Teilbaum. (Sie kennen das ja schon aus Übung 5, Aufgabe 2.) Sobald ein Ausdruck in Form eines Binärbaums im Hauptspeicher vorliegt, ist es einfach, diesen mittels Baumdurchlauf (in-, pre- oder postorder), wieder in eine Textform (z. B. In-, Prä- oder Postfix-Notation) zu übersetzen.

Die Repräsentation von arithmetischen Ausdrücken in Form von Binärbäumen bietet aber auch die Möglichkeit, einfache Optimierungen in den MidiPascal-Compiler einzubauen.

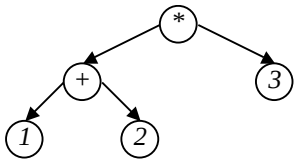
a) Ändern Sie die Erkennungsprozeduren für arithmetische Ausdrücke (*Expr*, *Term* und *Fact*) im Parser Ihres MidiPascal-Compilers so ab, dass vorerst kein Code mehr für die Ausdrücke erzeugt, sondern ein Binärbaum aufgebaut wird, dessen Knoten Zeichenketten enthalten (die vier Operatoren, die Ziffernfolge einer Zahl oder den Bezeichner einer Variablen).

b) Erweitern Sie das Code-Generierungsmodul dann um eine

```
PROCEDURE EmitCodeForExprTree(t: Tree);
```

die aus dem Binärbaum in einem Postorder-Durchlauf Code für die Berechnung des Ausdrucks durch die virtuelle MiniPascal-Maschine erzeugt.

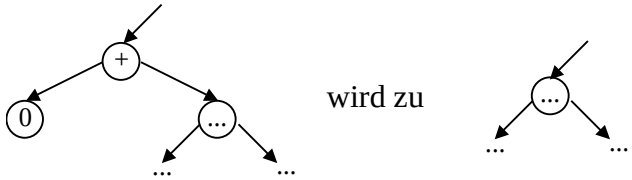
Beispiel: Für den Ausdruck (1 + 2) * 3 soll der links dargestellte Baum aufgebaut werden, und die Prozedur *EmitCodeForExprTree* soll daraus die rechts angegebene Codesequenz erzeugen:



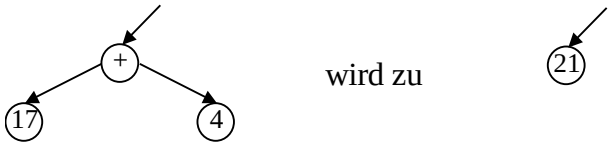

```
LoadConst 1
LoadConst 2
Add
LoadConst 3
Mul
```

Damit können Sie Ihren Compiler zwar schon testen – aber von Optimierung ist noch keine Rede. Die erzeugten Binärbäume eignen sich aber dazu, einfache Optimierungen an Ausdrücken vorzunehmen, die z. B. in modernen Compilern eingesetzt werden: die Binärbäume werden transformiert und erst die sich daraus ergebenden Bäume werden für die Codegenerierung herangezogen.

c) Eliminieren überflüssiger Rechenoperationen,
z. B.: 0 + ... oder ... + 0 oder 1 * ... oder ... * 1 oder ... / 1 wird zu ...
oder in Baumform (für das erste Beispiel) dargestellt:



d) „Konstantenfaltung“, Berechnung konstanter Teilausdrücke,
z. B.: ... + 17 + 4 + ... wird zu ... + 21 + ...



Versuchen Sie, möglichst viele solcher optimierender Baumtransformationen zu implementieren und wenden Sie diese solange auf den Baum an, als sich dadurch Verbesserungen ergeben.

Durch diese Transformationen sollte z. B. aus dem Baum für 0 + (17 + 4) * 1 ein Baum mit nur mehr einem Knoten für 21 entstehen.