Übersetzerbau VU
Übungsskriptum

Anton Ertl
Andreas Krall

2026

- Allgemeines und Beispiele
- GNU Emacs Reference Card
- AMD64-Assembler Handbuch
- make: A Program for Maintaining Programs
- lex — a Lexical Analyzer Generator
- yacc — Yet Another Compiler-Compiler
- Ox: Tutorial Introduction
- `burg`, `iburg` und `bfe`

1 Anmeldung

Melden Sie sich in TISS für die Lehrveranstaltung an. Nach Ende der Anmeldefrist (siehe Homepage) wird ein Account für Sie auf unserer Übungsmaschine `g0.complang.tuwien.ac.at` eingerichtet, der Accountname ist `u` gefolgt von der Matrikelnummer, z.B. u99999999. Ein Passwort für diesen Account erhalten Sie per Email. Bitte ändern Sie das Passwort möglichst bald, das zugesandte wird aus Sicherheitsgründen nach kurzer Zeit gesperrt.

2 Rechner

Die Übungsmaschine ist `g0.complang.tuwien.ac.at`; sollte sie längerfristig ausfallen, steht als Ersatzmaschine `g2.complang.tuwien.ac.at` zur Verfügung (Sie können sich aber vorerst nicht auf die Ersatzmaschine einloggen). Sie können sich aus dem Internet mit `ssh g0.complang.tuwien.ac.at` einloggen; falls Sie sich von einem Windows-Client aus einloggen wollen, können Sie das mit Putty tun.

Falls Sie Ihre Programme woanders entwickeln, müssen Sie Ihre Dateien für die Abgabe mit `scp` (eine ssh-Anwendung) auf unsere Rechner übertragen. Rufen Sie dann unbedingt das Test-Skript für das Beispiel *auf der Übungsmaschine* auf, damit Sie eventuelle Fehler mit katastrophalen Auswirkungen bemerken (wie z.B. das Kopieren in das falsche Verzeichnis).

Die in der Übung verwendeten Werkzeuge sind für verschiedene Plattformen auf `http://www.complang.tuwien.ac.at/ubvl/tools/` erhältlich (allerdings recht veraltet).

Wenn Sie selbst ein `.forward`-File einrichten oder ändern, testen Sie es unbedingt! Wenn es nicht funktioniert, haben wir keine Möglichkeit, Sie zu erreichen (z.B. um Ihnen die Ergebnisse der Abgabe mitzuteilen).

3 Betreuung, Information

Im WWW finden Sie unter `http://www.complang.tuwien.ac.at/ubvl/` Informationen zur Übung.

Verlautbarungen zur Übung (z.B. Klarstellungen zur Angabe) gibt es im Forum (Details dazu siehe Übungshomepage).

Wenn Sie eine Frage zur Übung haben, stellen Sie sie am besten im Forum (dann können auch andere antworten oder von der Antwort profitieren). Sie können auch den Leiter der Übung per Email fragen `anton@mips.complang.tuwien.ac.at`, oder in einer Sprechstunde (Termin per Email vereinbaren).

Technische Probleme wie Computerabstürze, falsche Permissions, oder vergessene Passwörter sind eine Sache für den Techniker. Wenden Sie sich di-

rekt an ihn: email an Herbert Pohlai (`herbert@mips.complang.tuwien.ac.at`), Tel. (+43-1) 58801/18525.

4 Beispiele

Die Beispiele finden Sie weiter hinten im Skriptum. Beachten Sie, dass die ersten Beispiele erfahrungsgemäß wesentlich leichter sind als die Beispiele „Attributierte Grammatik“ bis „Gesamtbeispiel“. Versuchen Sie, mit den ersten Beispielen möglichst rasch fertig zu werden, um genügend Zeit für die schwierigeren zu haben.

5 Beurteilung

Ihre Note wird aufgrund der Qualität der von Ihnen abgegebenen Programme ermittelt. Das Hauptkriterium ist dabei die Korrektheit. Sie wird mechanisch überprüft, Sie erhalten per Email das Ergebnis der Prüfung. Wenn Sie meinen, dass sich das Prüfprogramm geirrt hat, wenden Sie sich an den Leiter der Übung.

Die Prüfprogramme sind relativ einfach, dumm und kaum fehlertolerant. Damit Sie prüfen können, ob Ihr Programm im richtigen Format ausgibt und ähnliche wichtige Kleinigkeiten, stehen Ihnen die Testprogramme und einige einfache Testeingaben und -resultate zur Verfügung; Sie können die Testprogramme auch benutzen, um Ihre Programme mit eigenen Testfällen zu prüfen (siehe `http://www.complang.tuwien.ac.at/ubvl/`).

Beachten Sie, dass bei der Abgabe die Überprüfung mit wesentlich komplizierteren Testfällen erfolgt als denen, die wir Ihnen vorher zur Verfügung stellen (vor allem ab dem Scanner-Beispiel). Ein erfolgreiches Absolvieren der Ihnen vorher zur Verfügung stehenden Tests heißt also noch lange nicht, dass Ihr Programm korrekt ist. Sie müssen sich selbst weitere Testfälle überlegen (wie auch im Berufsleben).

Ihre Programme werden zu den angegebenen Terminen kopiert und später überprüft. Ändern Sie zu den Abgabeterminen zwischen 14h und 15h nichts im Abgabeverzeichnis, damit es nicht zu inkonsistenten Abgaben kommt.

Ein paar Tage nach der Abgabe erhalten Sie das Ergebnis per Email. Das Ausschicken der Ergebnisse wird auch im LVA-Forum verkündet, Sie brauchen also nicht nachfragen, wenn Sie dort noch nichts gesehen haben. Eine Arbeitswoche nach der ersten Abgabe werden Ihre (eventuell von Ihnen verbesserten) Programme erneut kopiert und überprüft. Diese Version wird mit 70% der Punkte eines rechtzeitig abgegebenen Programms gewertet. Das ganze wiederholt sich zwei Arbeitswochen nach dem ersten Abgabetermin (30% der Punkte). Sie erhalten für das Beispiel das Maximum der drei Ergebnisse.

Name	online Doku	Bemerkung
emacs, vi, code	info emacs, man vi	Editor
gcc	info as	Assembler
gcc	info gcc	C-Compiler
make	info make	baut Programme
flex	info flex	Scanner-Generator
bison, yacc	info bison, man yacc	Parser-Generator
dotty	`https://www.graphviz.org/Documentation.php`	Graphenzeichnen
ox, ox-1.04	`https://www.complang.tuwien.ac.at/ubvl/tools/doc/`	AG-basierter Compilergenerator
iburg, burg	`https://www.complang.tuwien.ac.at/ubvl/tools/doc/burg.pdf`	Baumparser-Generator
bfe	Skriptum	Präprozessor für burg
gdb	info gdb	Debugger
objdump	info objdump	Disassembler etc.

Abbildung 1: Werkzeuge

Sollten Sie versuchen, durch Kopieren oder Abschreiben von Programmen eine Leistung vorzutäuschen, die Sie nicht erbracht haben, erhalten Sie keine positive Note. Die Kontrolle erfolgt in einem Gespräch am Ende des Semesters, in dem überprüft wird, ob Sie auch verstehen, was Sie abgegeben haben.

Ihr Account ist nur für Sie lesbar. Bringen Sie andere nicht durch Ändern der Permissions in Versuchung, zu schummeln.

6 Weitere Dokumentation bzw. Werkzeuge

Abbildung 1 zeigt die zur Verfügung stehenden Werkzeuge.

Die mit „info“ gekennzeichnete Dokumentation können Sie mit dem Programm `info` lesen, oder indem sie in Emacs `C-h i` tippen. In der Dokumentation für Emacs bedeutet `C-`x Ctrl x und `M-`x Meta x (auf den Übungsgeräten also Alt x). Neben den genannten Dokumentationsquellen erfahren Sie mit `man` *`P`* für die meisten Programme, wie Sie *`P`* aufrufen können.

Mit flex erzeugte Scanner müssen normalerweise mit `-lfl` gelinkt werden.

Das auf den Übungsgeräten unter `yacc` aufrufbare Programm ist `bison -y`. Mit `dotty` können Sie sich die Ausgabe von `bison -g` anschauen. `ox-1.04` ist eine ältere Version von `ox`; sollte `ox` nicht so funktionieren wie dokumentiert, probieren Sie, ob `ox-1.04` stattdessen funktioniert (und wenn ja, ist Ihr Code interessant, um einen Bug report an den aktuellen Maintainer von Ox zu schicken).

7 Beispiele

Es sind insgesamt acht Beispiele abzugeben. Die ersten beiden Beispiele dienen dem Erlernen einiger grundlegender Befehle der AMD64-Architektur. In den weiteren Beispielen wird eine Programmiersprache vollständig implementiert. Diese Beispiele bauen aufeinander auf, d.h. Fehler, die Sie in den ersten Sprachimplementierungsbeispielen machen, sollten Sie beheben, damit sie in späteren Abgaben die Beurteilung nicht verschlechtern. Bei der Implementierung der Sprache wird mit jedem Beispiel (ausgenommen die letzten) auch ein neues Werkzeug eingeführt, das nach Einarbeitung in die Verwendungsweise des Werkzeugs die Arbeit erleichtert.

Die zu implementierende Sprache ist eingeschränkt, um den Arbeitsaufwand nicht zu groß werden zu lassen. So sind in dieser Sprache zwar grundlegende Kontrollstrukturen vorhanden und es können Variablen definiert werden, aber Datenstrukturen können innerhalb dieser Sprache nicht erzeugt werden. Sprachkonstrukte für Speicherzugriff und Adressarithmetik sind jedoch vorhanden. Daher müssen bei den letzten Beispielen, um die Codegenerierung testen zu können, Datenstrukturen in einem C-Programm erzeugt werden und dann mit dem von Ihnen generierten Code gelinkt werden. Dadurch erlernen Sie auch, wie verschiedene Sprachen miteinander kombiniert werden können. Weiters gibt es keine direkte Möglichkeit, Daten ein- oder auszugeben; diese Funktionen werden durch eine C-Funktion übernommen, die Funktionen in der Sprache aufruft, oder durch Aufrufen von C-Funktionen aus Funktionen in unserer Sprache.

Die Kenntnisse, die Sie bei den Assembler-Beispielen erlangen, werden Sie auch wieder bei der Codegenerierung der letzten Beispiele verwenden. Die Beispiele 3-8 können alle aufeinander aufbauend implementiert werden, d.h. wenn Sie Ihr Programm von Anfang an gut entwerfen, können Sie dieses ab dem Scanner-Beispiel bis zum Gesamtbeispiel stets wiederverwenden und erweitern. Beachten Sie jedoch, dass bei jeder Abgabe stets das gesamte Quellprogramm im Abgabeverzeichnis vorhanden sein muss (und zwar nicht in Form von symbolic links).

In den folgenden Abschnitten finden Sie die Angaben und Erklärungen für die Modalitäten der Beispielabgaben. Von der Sprache wird in jedem Abschnitt immer nur soviel erklärt, wie für das jeweilige Beispiel notwendig ist. Wenn Sie einen Überblick über die gesamte Sprache haben wollen, sollten Sie sich gleich am Anfang alle Angaben durchlesen.

In dieser Sprache kann man, wie in den meisten Programmiersprachen, auch Programme schreiben, deren Semantik nicht definiert ist, und die Ihr Compiler trotzdem nicht als fehlerhaft erkennen muss und darf. Bei solchen Programmen ist es egal, welchen Code Ihr Compiler produziert (Code aus solchen Testeingaben wird von unseren Abgabescripts ohnehin nicht ausgeführt). Ihr Compiler sollte aber für Programme mit definierter Semantik

korrekten Code produzieren.

7.1 Assembler A

7.1.1 Termin

Abgabe spätestens am 18. März 2026, 14 Uhr.

7.1.2 Angabe

Die Berechnung

$$r = mn + o$$

wobei r, m und o aus mehreren 64b-Maschinenwörtern bestehen und n aus einem, kann aus Berechnungen zusammengesetzt werden, in denen nur ein oder zwei Maschinenwörter (64b bzw. 128b) vorkommen:

$$c_{i+1}b + r_i = (m_i n + c_i) + o_i$$

wobei die Multiplikation hier 64b-Operanden, aber ein 128b-Resultat hat (hier dargestellt durch zwei 64b-Teile; b ist 2^{64}), und die Additionen dann jeweils einen 128b-Operanden und einen 64b-Operanden haben (alle jeweils vorzeichenlos). r_i ist ein Glied (64b-Zahl) des Ergebnisses. c_i ist der Übertrag (eine 64b-Zahl) vom vorigen Schritt, und c_{i+1} der Übertrag zum nächsten. Wenn der aktuelle Schritt der letzte ist, ist das höchstwertige Wort des Resultats $r_{i+1} = c_{i+1}$.

Für die 128b+64b-Addition gibt es keinen Befehl auf AMD64, aber man kann den ganzen Schritt auch so formulieren:

$$(d_{i+1} + C_{i+1} + O_{i+1})b + r_i = (m_i n + d_i + C_i) + o_i + O_i$$

Hier ist $d_{i+1} = (m_i n)[127 : 64]$, wobei die Notation $x[127 : 64]$ die Bits 64 bis 127 von x bedeuten. C_{i+1} ist das carry-bit (Bit 64) von $s = (m_i n)[63 : 0] + d_i + C_i$; O_{i+1} ist das carry-bit (Bit 64) von $t = s[63 : 0] + o_i + O_i$, während $r_i = t[63 : 0]$. Die Summe $d_{i+1} + C_{i+1} + O_{i+1}$ hat den selben Wert wie c_{i+1} in der obigen Berechnung, und entsprechend kommt aus dem vorigen Schritt $d_i + C_i + O_i$. Um am Schluss die höchstwertigen 64b von r zu berechnen, muss man die drei Teile zusammenaddieren:

$$r_n = d_n + C_n + O_n$$

Für die Addition zweier 64b-Zahlen mit carry-in und carry-out hat AMD64 Befehle: `adcx` und `adox`.

Schreiben Sie unter Verwendung der Befehle `adcx` und `adox` eine Funktion mit dem Prototypen:

```
void asma(unsigned long *m, unsigned long n, unsigned long *o,
          unsigned long *r);
```

die eine 128b-Zahl bei `m` mit `n` multipliziert und die 128b-Zahl bei `o` dazuaddiert; das Ergebnis ist die 192b-Zahl bei `r`.

Dabei ist der Befehl `mulx` hilfreich, der die Flags nicht zerstört und mehr Flexibilität bei den Operanden bietet als `mul`.

Am einfachsten tun Sie sich dabei wahrscheinlich, wenn Sie eine ähnliche Funktion in C schreiben und mit z.B. `gcc -O -S` in Assembler übersetzen und das Ergebnis dann verändern. Dann stimmt schon das ganze Drumherum.

7.1.3 Hinweis

Beachten Sie, dass Sie nur dann Punkte bekommen, wenn Ihre Version die verlangten Befehle ausführt und korrekt ist.

Zum Assemblieren und Linken verwendet man am besten `gcc`, der Compiler-Treiber kümmert sich dann um die richtigen Optionen für `as` und `ld`.

7.1.4 Abgabe

Zum angegebenen Termin stehen im Verzeichnis `~/abgabe/asma` die maßgeblichen Dateien. Mittels `make clean` soll man alle von Werkzeugen erzeugten Dateien löschen können und `make` soll eine Datei `asma.o` erzeugen. Diese Datei soll nur die Funktion `asma` enthalten, keinesfalls `main`. Diese Funktion soll den Aufrufkonventionen gehorchen und wird bei der Prüfung der abgegebenen Programme mit C-Code zusammengebunden.

7.2 Assembler B

7.2.1 Termin

Abgabe spätestens am 25. März 2026, 14 Uhr.

7.2.2 Angabe

Gegeben ist folgender Funktionsprototyp:

```
void asmb(unsigned long *m, unsigned long n, unsigned long *o,
          unsigned long *r, size_t n);
```

Implementieren Sie eine Funktion mit diesem Prototyp in Assembler, die eine $64n$-bit-Zahl bei `m` mit `n` multipliziert und die $64n$-bit-Zahl bei `o` dazuaddiert. Das Ergebnis ist eine $64(n+1)$-bit-Zahl bei `r`.

Verwenden Sie dazu die Befehle `adcx` und `adox`; weiters können `mulx` und `loop` hilfreich sein, weil diese Befehle die Flags nicht verändern.

Für besonders effiziente Lösungen (gemessen an der Anzahl der *ausgeführten* Maschinenbefehle; wird ein Befehl n mal ausgeführt, zählt er n-fach) gibt es Bonuspunkte.

7.2.3 Hinweis

Beachten Sie, dass Sie nur dann Punkte bekommen, wenn Ihre Funktion die verlangten Befehle ausführt und korrekt ist. Dadurch können Sie viel mehr verlieren als Sie durch Optimierung gewinnen können, also optimieren Sie im Zweifelsfall lieber weniger als mehr.

Die Vertrautheit mit dem Assembler müssen Sie beim Gespräch am Ende des Semesters beweisen, indem Sie Fragen zum abgegebenen Code beantworten.

7.2.4 Abgabe

Zum angegebenen Termin stehen im Verzeichnis `~/abgabe/asmb` die maßgeblichen Dateien. Mittels `make clean` soll man alle von Werkzeugen erzeugten Dateien löschen können und `make` soll eine Datei `asmb.o` erzeugen. Diese Datei soll nur die Funktion `asmb` enthalten, keinesfalls `main`. Diese Funktion soll den Aufrufkonventionen gehorchen und wird bei der Prüfung der abgegebenen Programme mit C-Code zusammengebunden.

7.3 Scanner

7.3.1 Termin

Abgabe spätestens am 15. April 2026, 14 Uhr.

7.3.2 Angabe

Schreiben Sie mit `flex` einen Scanner, der Identifier, Zahlen, und folgende Schlüsselwörter unterscheiden kann: `end return var cond continue break and not`. Weiters soll er auch noch folgende Lexeme erkennen: `; ( ) , := : -> * + > = - [ ]`

Identifier bestehen aus Buchstaben, Ziffern, und `_`, und müssen mit einem Buchstaben beginnen. Zahlen sind entweder Hexadezimalzahlen oder Dezimalzahlen. Hexadezimalzahlen beginnen mit `0x`, gefolgt von einer oder mehr Hexadezimalziffern oder `_`, wobei Hex-Ziffern sowohl groß als auch klein geschrieben sein dürfen. Dezimalzahlen bestehen aus Dezimalziffern und `_`, wobei mindestens eine Dezimalziffer vorkommen muss. Leerzeichen, Tabs und Newlines zwischen den Lexemen sind erlaubt und werden ignoriert, ebenso Kommentare, die mit `-` anfangen und bis zum Ende der Zeile gehen). Alles andere sind lexikalische Fehler. Es soll jeweils das längste mögliche Lexem

erkannt werden, `end39` ist also ein Identifier (longest input match), `39end` ist die Zahl `39` gefolgt vom Schlüsselwort `end`.

Der Scanner soll für jedes Lexem eine Zeile ausgeben: für Schlüsselwörter und Lexeme aus Sonderzeichen soll das Lexem ausgegeben werden, für Identifier `id` gefolgt von einem Leerzeichen und dem String des Identifiers, für Zahlen `num` gefolgt von einem Leerzeichen und der Zahl in Dezimaldarstellung ohne führende Nullen. Für Leerzeichen, Tabs, Newlines und Kommentare soll nichts ausgegeben werden (auch keine Leerzeile).

Der Scanner soll zwischen Groß- und Kleinbuchstaben unterscheiden, `End` ist also kein Schlüsselwort.

7.3.3 Abgabe

Legen Sie ein Verzeichnis `~/abgabe/scanner` an, in das Sie die maßgeblichen Dateien stellen. Mittels `make clean` soll man alle von Werkzeugen erzeugten Dateien löschen können (auch den ausführbaren Scanner) und mittels `make` ein Programm namens `scanner` erzeugen, das von der Standardeingabe liest und auf die Standardausgabe ausgibt. Korrekte Eingaben sollen akzeptiert werden (Ausstieg mit Status 0, z.B. mit `exit(0)`), bei einem lexikalischen Fehler soll der Fehlerstatus 1 erzeugt werden. Bei einem lexikalischen Fehler darf der Scanner Beliebiges ausgeben (eine sinnvolle Fehlermeldung hilft bei der Fehlersuche).

7.3.4 Hinweis

Die lex-Notation `$` steht für ein ein Zeilenende, auf das ein Newline folgt; zusätzlich kann auch noch das Ende der Eingabe die Zeile (und damit einen Kommentar) beenden. Am einfachsten ist es, nur zu spezifizieren, was ein Kommentar ist, und es dem longest input match zu überlassen, den Kommentar nicht zu früh abzubrechen.

7.4 Parser

7.4.1 Termin

Abgabe spätestens am 22. April 2026, 14 Uhr.

Abbildung 2 zeigt die Grammatik (in `yacc`/`bison`-artiger EBNF) einer Programmiersprache. Schreiben Sie einen Parser für diese Sprache mit `flex` und `yacc`/`bison`. Die Lexeme sind die gleichen wie im Scanner-Beispiel (`id` steht für einen Identifier, `num` für eine Zahl). Das Startsymbol ist `Program`.

```
Program: { Funcdef ';' }
       ;

Funcdef: id '(' Pars ')' Stats end  /* Funktionsdefinition */
       ;

Pars: { id ',' } [ id ]     /* Parameterdefinition */
    ;

Stats: { Stat ';' }
     ;

Stat: return Expr
    | Conds
    | var id ':=' Expr /* Variablendefinition */
    | Lexpr ':=' Expr  /* Zuweisung */
    | Term
    ;

Conds: [ id ':' ] /* Labeldefinition */
        cond { Guarded ';' } end
      ;

Guarded: [ Expr ] '->' Stats
         ( continue | break ) [ id ] /* Labelverwendung */
       ;

Lexpr: id        /* schreibender Variablenzugriff */
     | Term '[' Expr ']' /* schreibender Arrayzugriff */
     ;

Expr: { not } Term
    | Term '[' Expr ']'   /* lesender Arrayzugriff */
    | Term { '+' Term }
    | Term { '*' Term }
    | Term { and Term }
    | Term ( '>' | '=' | '-' ) Term
    ;

Term: '(' Expr ')'
    | num
    | id                                /* Variablenverwendung */
    | id '(' { Expr ',' } [ Expr ] ')' /* Funktionsaufruf */
    ;
```

Abbildung 2: Grammatik

7.4.2 Abgabe

Zum angegebenen Termin stehen im Verzeichnis `~/abgabe/parser` die maßgeblichen Dateien. Mittels `make clean` soll man alle von Werkzeugen erzeugten Dateien löschen können und mittels `make` ein Programm namens `parser` erzeugen, das von der Standardeingabe liest. Korrekte Programme sollen akzeptiert werden (Ausstieg mit Status 0, z.B. mit `exit(0)`), bei einem lexikalischen Fehler soll der Fehlerstatus 1 erzeugt werden, bei Syntaxfehlern der Fehlerstatus 2. Das Programm darf auch etwas ausgeben (auch bei korrekter Eingabe), z.B. damit Sie sich beim Debugging leichter tun.

7.4.3 Hinweis

Die Verwendung von Präzedenzdeklarationen von `yacc` kann leicht zu Fehlern führen, die man nicht so schnell bemerkt (bei dieser Grammatik sind sie sowieso sinnlos). Konflikte in der Grammatik sollten Sie durch Umformen der Grammatik beseitigen; `yacc` löst den Konflikt zwar, aber nicht unbedingt in der von Ihnen gewünschten Art.

Links- oder Rechtsrekursion? Also: Soll das rekursive Vorkommen eines Nonterminals als erstes (links) oder als letztes (rechts) auf der rechten Seite der Regel stehen? Bei `yacc`/`bison` und anderen LR-basierten Parsergeneratoren funktioniert beides. Sie sollten sich daher in erster Linie danach richten, was leichter geht, z.B. weil es Konflikte vermeidet oder weil es einfachere Attributierungsregeln erlaubt. Z.B. kann man mittels Linksrekursion bei der Subtraktion einen Parse-Baum erzeugen, der auch dem Auswertungsbaum entspricht.[1] Sollte es keine anderen Gründe geben, kann man der Linksrekursion den Vorzug geben, weil sie mit einer konstanten Tiefe des Parser-Stacks auskommt.

7.5 Attributierte Grammatik

7.5.1 Termin

Abgabe spätestens am 6. Mai 2026, 14 Uhr.

7.5.2 Angabe

Erweitern Sie den Parser aus dem letzten Beispiel mit Hilfe von `ox` um eine Symboltabelle und eine statische Analyse.

Die *hervorgehobenen* Begriffe beziehen sich auf Kommentare in der Grammatik.

[1] Das betrifft nicht die Grammatik in dieser Übung; in dieser Grammatik darf man nicht `x-y-z` schreiben, sondern muss durch Klammern klarmachen, ob `(x-y)-z` oder `x-(y-z)` gemeint ist (die übliche Interpretation ist `(x-y)-z`.

7.5.2.1 Namen. Die folgenden Dinge haben Namen: Funktionen, Variablen, und Labels.

An einer Stelle darf nur eine Variable oder ein Label mit dem selben Namen sichtbar sein.

Alle Namen (`ids`), die in einer *Parameterdefinition* oder in einer *Variablendefinition* vorkommen, sind Variablennamen. Variablen, die in einer Parameterdefinition definiert wurden, sind in der ganzen Funktion sichtbar. Variablen, die einer Variablendefinition definiert wurden, sind in allen folgenden Statements der unmittelbar umgebenden `Stats` sichtbar, und nirgendwo sonst. In der Definition selbst ist die Variable noch nicht sichtbar.

Bei einer *Variablenverwendung* und einem *schreibenden Variablenzugriff* muss eine Variable mit dem Namen sichtbar sein.

In einer *Labeldefinition* vorkommende Namen sind Labels. Ein Label ist im gesamten `Conds` sichtbar, vor dem seine Definition steht. Bei einer Labelverwendung muss ein Label mit dem Namen sichtbar sein.

Eine Funktion wird im *Funktionsaufruf* verwendet und in der *Funktionsdefinition* definiert. Verwendete Funktionen müssen nicht definiert werden und können nicht deklariert[2] werden. Funktionen dürfen, soweit es den Compiler betrifft, doppelt definiert werden und dürfen den gleichen Namen wie Variablen oder Labels haben; daher muss der Compiler Funktionsnamen nicht in einer Symboltabelle verwalten. Auch die Übereinstimmung der Anzahl der Argumente soll (und, im allgemeinen, kann) der Compiler nicht überprüfen.

Ihre statisch Analyse soll überprüfen, dass alle verwendeten Variablen- und Label-Namen sichtbar sind und der richtigen Klasse (Label oder Variable) angehören, und dass zwei Definitionen des gleichen Label- oder Variablen-Namens keine überlappenden Sichtbarkeitsbereiche haben (auch nicht mit einem Namen der anderen Klasse).

7.5.3 Hinweise

Es ist empfehlenswert, die Grammatik so umzuformen, dass sie für die AG günstig ist: Fälle, die syntaktisch gleich ausschauen, aber bei den Attributierungsregeln verschieden behandelt werden müssen, sollten auf verschiedene Regeln aufgeteilt werden; umgekehrt sollten Duplizierungen, die in dem Bemühen vorgenommen wurden, Konflikte zu vermeiden, auf ihre Sinnhaftigkeit überprüft und ggf. rückgängig gemacht werden. Testen Sie Ihre Grammatikumformungen mit den Testfällen.

Offenbar übersehen viele Leute, dass attributierte Grammatiken Information auch von rechts nach links (im Ableitungsbaum) weitergeben können.

[2] Im Sinne von C: Die Definition einer Funktion enthält den vollständigen Code. Die Deklaration enthält nur die Informationen, die der Compiler braucht, um eine Typüberprüfung des Aufrufs durchzuführen (in C auch bekannt als Prototyp, in anderen Sprachen oft als Signatur).

Sie denken sich dann recht komplizierte Lösungen aus. Dabei reichen die von `ox` zur Verfügung gestellten Möglichkeiten vollkommen aus, um zu einer relativ einfachen Lösung zu kommen. Heuer sind diese Möglichkeiten zwar für das AG-Beispiel wohl nicht nötig, aber behalten Sie sie für spätere Beispiele im Hinterkopf.

Verwenden Sie keine globalen Variablen oder Funktionen mit Seiteneffekten (z.B. Funktionen, die übergebene Datenstrukturen ändern) bei der Attributberechnung! `ox` macht globale Variablen einerseits unnötig, andererseits auch fast unbenutzbar, da die Ausführungsreihenfolge der Attributberechnung nicht vollständig festgelegt ist. Bei Traversals ist die Reihenfolge festgelegt, und Sie können globale Variablen verwenden; seien Sie aber trotzdem vorsichtig.

Sie brauchen angeforderten Speicher (z.B. für Symboltabellen-Einträge oder Typinformation) nicht freigeben, die Testprogramme sind nicht so groß, dass der Speicher ausgeht (zumindest wenn Sie's nicht übertreiben).

Das Werkzeug Torero (`http://www.complang.tuwien.ac.at/torero/`) ist dazu gedacht, bei der Erstellung von attributierten Grammatiken zu helfen.

7.5.4 Abgabe

Zum angegebenen Termin stehen die maßgeblichen Dateien im Verzeichnis `~/abgabe/ag`. Mittels `make clean` soll man alle von Werkzeugen erzeugten Dateien löschen können und mittels `make` ein Programm namens `ag` erzeugen, das von der Standardeingabe liest. Korrekte Programme sollen akzeptiert werden, bei einem lexikalischen Fehler soll der Fehlerstatus 1 erzeugt werden, bei Syntaxfehlern der Fehlerstatus 2, bei anderen Fehlern (z.B. Verwendung eines nicht sichtbaren Namens) der Fehlerstatus 3. Die Ausgabe kann beliebig sein, auch bei korrekter Eingabe.

7.6 Codeerzeugung A

7.6.1 Termin

Abgabe spätestens am 20. Mai 2026, 14 Uhr.

7.6.2 Angabe

Erweitern Sie die statische Analyse aus dem AG-Beispiel mit Hilfe von `iburg` zu einem Compiler, der folgende Untermenge der statisch korrekten Programme in AMD64-Assemblercode übersetzt: alle Programme, in denen aus `Stat` nur `return`-Anweisungen abgeleitet werden, in denen aber kein *Funktionsaufruf* abgeleitet wird. Programme, die statisch korrekt sind, aber dieser

Einschränkung nicht entsprechen, werden bei diesem Beispiel nicht als Testeingaben vorkommen.

Ein Teil der Sprache wurde schon im Beispiel attributierte Grammatik erklärt, hier der für dieses Beispiel notwendige Zusatz:

7.6.2.1 Datendarstellung. Diese Programmiersprache kennt nur einen Datentyp: das 64-bit-Wort, das als vorzeichenbehaftete Zahl oder als Arrayadresse verwendet werden kann. Weder der Compiler noch das Laufzeitsystem soll eine Typüberprüfung vornehmen. Der Programmierer (der Anwender des Compilers) muss wissen, was er tut, der Compiler soll (und kann) das nicht überprüfen. Unsere Testprogramme führen keine Zugriffe auf ungültige Adressen aus.

7.6.2.2 Bedeutung der Operatoren. `+`, `*` und `-` haben ihre übliche Bedeutung (ein etwaiger Überlauf soll ignoriert werden). `and` und `not` führen ihre Operationen bitweise auf ihren Operanden aus. `>` und `=` vergleichen ihre Operanden und liefern 1 für wahr und 0 für falsch.

Bei einem (schreibenden oder lesenden) *Arrayzugriff* ist der vom `Term` (auf der rechten Seite der Regel) berechnete Wert die Addresse von Element 0 des Arrays, und der von `Expr` berechnete Wert der Index des Elements, auf das zugegriffen wird. Der Zugriff erfolgt dann auf die Adresse Term+8*Expr. Der *lesende Arrayzugriff* liefert als Resultat den 64-bit-Wert an dieser Adresse.

7.6.2.3 Anweisungen Die `return`-Anweisung beendet die Funktion und liefert das Resultat von `Expr` als Ergebnis des Aufrufs der Funktion.

7.6.2.4 Erzeugter Code. Ihr Compiler soll AMD64-Assemblercode ausgeben. Jede Funktion im Programm verhält sich gemäß der Aufrufkonvention. Der erzeugte Code wird nach dem Assemblieren und Linken von C-Funktionen aufgerufen. Beispiel: Die Funktion `foo(a,b) ... end;` kann von C aus mit `foo(x,y)` aufgerufen werden, wobei `a` den Wert von `x` bekommt und `b` den von `y`.

Der Name einer Funktion soll als Assembler-Label am Anfang des erzeugten Codes verwendet werden und das Symbol soll exportiert werden; andere Symbole soll Ihr Code nicht exportieren.

Folgende Einschränkungen sind dazu gedacht, Ihnen gewisse Probleme zu ersparen, die reale Compiler bei der Codeauswahl und Registerbelegung haben. Sie brauchen diese Einschränkungen nicht überprüfen, unsere Testeingaben halten sich an diese Einschränkungen (eine Überprüfung könnte Ihnen allerdings beim Debuggen Ihrer eigenen Testeingaben helfen): Funktionen ha-

ben maximal 6 Parameter. Die maximale Tiefe eines Ausdrucks[3] ist $\leq 6 - v$, wobei v die Anzahl der sichtbaren Variablen ist. Die im Quellprogramm vorkommenden Zahlen und konstanten Ausdrücke sind $\geq -2^{31}$ und $< 2^{31}$; das gilt aber nicht für Ergebnisse von Berechnungen zur Laufzeit.

Der erzeugte Code soll korrekt sein und möglichst wenige Befehle ausführen (da es hier keine Verzweigungen gibt, ist das gleichbedeutend mit „wenige Befehle enthalten“). Dabei ist nicht an eine zusätzliche Optimierung (wie z.B. common subexpression elimination) gedacht, sondern vor allem an die Dinge, die Sie mit `iburg` tun können, also eine gute Codeauswahl (besonders bezüglich konstanter Operanden und Ausnutzung der Adressierungsarten) und eventuell einige algebraische Optimierungen (siehe z.B. `http://www.complang.tuwien.ac.at/papers/ertl00dagstuhl.ps.gz`). Für besonders effizienten erzeugten Code gibt es Sonderpunkte.

Beachten Sie, dass es leicht ist, durch eine falsche Optimierungsregel mehr Punkte zu verlieren, als Sie durch Optimierung überhaupt gewinnen können. Testen Sie daher ihre Optimierungen besonders gut (mindestens ein Testfall pro Optimierungsregel). Überlegen Sie sich, welche Optimierungen es wohl wirklich bringen (welche Fälle also tatsächlich vorkommen), und lassen Sie die anderen weg.

7.6.3 Abgabe

Zum angegebenen Termin stehen die maßgeblichen Dateien im Verzeichnis `~/abgabe/codea`. Mittels `make clean` soll man alle von Werkzeugen erzeugten Dateien löschen können und mittels `make` ein Programm namens `codea` erzeugen, das von der Standardeingabe liest und den generierten Code auf die Standardausgabe ausgibt. Bei einem lexikalischen Fehler soll der Fehlerstatus 1 erzeugt werden, bei einem Syntaxfehler Fehlerstatus 2, bei anderen Fehlern der Fehlerstatus 3. Im Fall eines Fehlers darf die Ausgabe beliebig sein.

7.7 Codeerzeugung B

7.7.1 Termin

Abgabe spätestens am 3. Juni 2026, 14 Uhr.

7.7.2 Angabe

Erweitern Sie den Compiler aus dem vorigen Beispiel so, dass er folgende Untermenge der statisch korrekten Programme in AMD64-Assemblercode übersetzt: Alle Programme, in denen der Parser keinen *Funktionsaufruf* ableitet.

[3]Tiefe eines Ausdrucks: Anzahl der Ableitungen von `Expr` zwischen einem Blatt des Syntaxbaums und dem nächsten `Statement`.

Programme, die statisch korrekt sind, aber dieser Einschränkung nicht entsprechen, werden bei diesem Beispiel nicht als Testeingaben vorkommen.

Ein Teil der Sprache wurde schon erklärt, hier der für dieses Beispiel notwendige Zusatz:

Die *Variablendefinition* speichert den Wert von `Expr` unter dem Namen der Variable.

Bei einer *Zuweisung* wird `Expr` ausgewertet und das Resultat in die Variable bzw. in das Arrayelement gespeichert.

Eine `Term`-Anweisung wertet den Term aus und macht mit dem Ergebnis nichts (in diesem Beispiel gibt es keine Funktionsaufrufe, daher macht diese Anweisung hier gar nichts).

Bei einem `Conds` wird das `Expr` des ersten enthaltenen `Guarded` ausgewertet; wenn das Resultat ungerade ist, werden die enthaltenen `Stats` ausgeführt, ansonsten wird mit dem nächsten `Guarded` weitergemacht. Wenn ein `Guarded` kein `Expr` vor dem `->` hat, werden die enthaltenen `Stats` auf jeden Fall ausgeführt. Wenn es kein nächstes `Guarded` gibt, wird hinter dem `Conds` weitergemacht. Wenn in einem `Guarded` das Ende der `Stats` erreicht wird, springt `continue` an den Anfang des vom Label `id` bezeichneten `Condss`, `break` an das Ende; ist kein Label angegeben, bezieht sich `continue` bzw. `break` auf das unmittelbar umgebende Conds.

7.7.2.1 Erzeugter Code. Es gelten die gleichen Anforderungen und Einschränkungen wie im vorigen Beispiel.

7.7.3 Hinweis

Es bringt nichts, für `iburg` Bäume zu bauen, die mehr als eine einfache Anweisung oder einen Vergleich umfassen: die Möglichkeit, durch die Baumgrammatik Knoten zusammenzufassen und so zu optimieren, kann nur auf der Ebene von Ausdrücken und einfachen Anweisungen genutzt werden (ausser man würde die Zwischendarstellung in einer Weise umformen, die zuviel Aufwand für diese LVA ist).

Auf höherer Ebene ist einfacher, für jede einfache Anweisung einen Baum zu bauen und dann in einem Traversal für jeden dieser Bäume den Labeler und den Reducer aufzurufen.

7.7.4 Abgabe

Zum angegebenen Termin stehen die maßgeblichen Dateien im Verzeichnis `~/abgabe/codeb`. Mittels `make clean` soll man alle von Werkzeugen erzeugten Dateien löschen können und mittels `make` ein Programm namens `codeb` erzeugen, das von der Standardeingabe liest und den generierten Code auf

die Standardausgabe ausgibt. Bei einem lexikalischen Fehler soll der Fehlerstatus 1 erzeugt werden, bei einem Syntaxfehler Fehlerstatus 2, bei anderen Fehlern der Fehlerstatus 3. Im Fall eines Fehlers darf die Ausgabe beliebig sein.

7.8 Gesamtbeispiel

7.8.1 Termin

Abgabe spätestens am 17. Juni 2026, 14 Uhr.

Es gibt nur einen Nachtermin.

7.8.2 Angabe

Erweitern Sie den Compiler aus dem vorigen Beispiel so, dass er alle statisch korrekten Programme in AMD64-Assemblercode übersetzt.

Ein Teil der Sprache wurde schon erklärt, hier der für dieses Beispiel notwendige Zusatz:

Der *Funktionsaufruf* wertet alle `Expr`s aus und ruft dann die Funktion `id` auf, mit den Ergebnissen der `Expr`s als Parameter. Der von der Funktion zurückgegebene Wert ist der Wert des Funktionsaufrufs.

7.8.2.1 Erzeugter Code. Der erzeugte Code ruft Funktionen entsprechend den Aufrufkonventionen auf. Ansonsten gelten die gleichen Anforderungen und Einschränkungen wie im vorigen Beispiel, wobei ein Funktionsaufruf mit n Parametern bei der Berechnung der Tiefe mit dem Wert $\max(0, n-1)$ (zuzüglich der maximalen Tiefe der Berechnungen der Parameter) eingeht.

Wichtigstes Kriterium ist wie immer die Korrektheit, für gute Codeerzeugung gibt es aber wieder Sonderpunkte. Wir empfehlen, nur Optimierungen durchzuführen, die mit den verwendeten Werkzeugen einfach möglich sind. Bei diesem Beispiel kommt es mehr auf gute Registerbelegung an als auf die Optimierung von Ausdrücken.

7.8.3 Hinweise

Bei der Registerbelegung gibt es sowohl ein großes Optimierungspotential als auch ein großes Fehlerpotential, besonders im Zusammenhang mit (verschachtelten) Funktionsaufrufen.

Eine einfache Strategie bezüglich der Parameter der aktuellen Funktion ist, sie nicht in den Argumentregistern zu lassen, sondern sie z.B. auf den Stack zu kopieren, damit man beim Berechnen der Parameter einer anderen Funktion problemlos auf sie zugreifen kann. Diese Strategie mag zwar nicht zum optimalen Code führen, aber eine gute Regel beim Programmieren lautet: “First make it work, then make it fast”.

7.8.4 Abgabe

Zum angegebenen Termin stehen die maßgeblichen Dateien im Verzeichnis `~/abgabe/gesamt`. Mittels `make clean` soll man alle von Werkzeugen erzeugten Dateien löschen können und mittels `make` ein Programm namens `gesamt` erzeugen, das von der Standardeingabe liest und auf die Standardausgabe ausgibt. Bei einem lexikalischen Fehler soll der Fehlerstatus 1 erzeugt werden, bei einem Syntaxfehler Fehlerstatus 2, bei anderen Fehlern der Fehlerstatus 3. Im Fall eines Fehlers kann die Ausgabe beliebig sein. Der ausgegebene Code muss vom Assembler verarbeitet werden können.