

Prinzipien von Programmiersprachen Vorlesung Wintersemester 2007 – 2020

Johannes Waldmann, HTWK Leipzig

26. Januar 2021

1 Einleitung

Programme und Algorithmen

- Algorithmus (vgl. VL Alg. und Datenstr.)
Vorschrift zur Lösung einer Aufgabe
- Programm (vgl. VL zu (Anwendungsorientierter) Progr.)
Realisierung eines Algorithmus in konkreter Programmiersprache, zur Ausführung durch Maschine
- Programmiersprache
bietet Ausdrucksmittel zur Realisierung von Algorithmen als Programme

Deutsch als Programmiersprache

§6 (2) ...Der Zuteilungsdivisor ist so zu bestimmen, dass insgesamt so viele Sitze auf die Landeslisten entfallen, wie Sitze zu vergeben sind. Dazu wird zunächst die Gesamtzahl der Zweitstimmen aller zu berücksichtigenden Landeslisten durch die Zahl der jeweils nach Absatz 1 Satz 3 verbleibenden Sitze geteilt. Entfallen danach mehr Sitze auf die Landeslisten, als Sitze zu vergeben sind,...

§6 (5) Die Zahl der nach Absatz 1 Satz 3 verbleibenden Sitze wird so lange erhöht, bis jede Partei bei der zweiten Verteilung der Sitze nach Absatz 6 Satz 1 mindestens die bei der ersten Verteilung nach den Absätzen 2 und 3 für sie ermittelten zuzüglich der in den Wahlkreisen errungenen Sitze erhält, die nicht nach Absatz 4 Satz 1 von der Zahl der für die Landesliste ermittelten Sitze abgerechnet werden können.

https://www.gesetze-im-internet.de/bwahlg/___6.html

Beispiel: mehrsprachige Projekte

ein typisches Projekt besteht aus:

- Datenbank: SQL
- Verarbeitung: Java
- Oberfläche: HTML
- Client-Code: Java-Script

und das ist noch nicht die ganze Wahrheit:

nenne weitere Sprachen, die üblicherweise in einem solchen Projekt vorkommen

In / Into

- David Gries (1981) zugeschrieben, zitiert u.a. in McConnell: Code Complete, 2004.
Unterscheide:
 - programming *in* a language
Einschränkung des Denkens auf die (mehr oder weniger zufällig) vorhandenen Ausdrucksmittel
 - programming *into* a language
Algorithmus → Programm
- Ludwig Wittgenstein: Die Grenzen meiner Sprache sind die Grenzen meiner Welt (sinngemäß — Ü: Original?)
- Folklore:
A good programmer can write LISP in any language.

Sprache

- wird benutzt, um Ideen festzuhalten/zu transportieren (Wort, Satz, Text, Kontext)
- wird beschrieben durch
 - Lexik
 - Syntax
 - Semantik
 - Pragmatik
- natürliche Sprachen / formale Sprachen

Wie unterschiedlich sind Sprachen?

- weitgehend übereinstimmende Konzepte.
 - LISP (1958) = Perl = PHP = Python = Ruby = Javascript = Clojure: imperativ, (funktional), nicht statisch typisiert (d.h., unsicher und ineffizient)
 - Algol (1958) = Pascal = C = Java = C#
imperativ, statisch typisiert
 - ML (1973) = Haskell:
statisch typisiert, generische Polymorphie
- echte Unterschiede („Neuerungen“) gibt es auch
 - CSP (1977) = Occam (1983) = Go: Prozesse, Kanäle
 - Clean (1987) \approx Rust (2012): Lineare Typen
 - Coq (1984) = Agda (1999) = Idris: dependent types

Konzepte

- Hierarchien (baumartige Strukturen)
 - einfache und zusammengesetzte (arithmetische, logische) Ausdrücke
 - einfache und zusammengesetzte Anweisungen (strukturierte Programme)
 - Komponenten (Klassen, Module, Pakete)
- Typen beschreiben Daten, gestatten statische Prüfung
- Namen stehen für Werte, gestatten Wiederverwendung
- flexible Wiederverwendung durch Parameter (Argumente)
Unterprogramme: Daten, Polymorphie: Typen

Paradigmen

- imperativ
Programm ist Folge von Befehlen
(Befehl bewirkt Zustandsänderung)
- deklarativ (Programm ist Spezifikation)
 - funktional (Gleichungssystem)
 - logisch (logische Formel über Termen)
 - Constraint (log. F. über anderen Bereichen)
- objektorientiert (klassen- oder prototyp-basiert)
- nebenläufig (nichtdeterministisch, explizite Prozesse)
- (hoch) parallel (deterministisch, implizit)

Ziele der LV

Arbeitsweise: Methoden, Konzepte, Paradigmen

- isoliert beschreiben
- an Beispielen in (bekannten und unbekanntem) Sprachen wiedererkennen

Ziel:

- verbessert die Organisation des vorhandenen Wissens
- gestattet die Beurteilung und das Erlernen neuer Sprachen
- hilft bei Entwurf eigener (anwendungsspezifischer) Sprachen

Beziehungen zu anderen LV

- Grundlagen der Informatik, der Programmierung:
strukturierte (imperative) Programmierung
- Softwaretechnik 1/2:
objektorientierte Modellierung und Programmierung, funktionale Programmierung
und OO-Entwurfsmuster

- Compilerbau: Implementierung von Syntax und Semantik

Sprachen für bestimmte Anwendungen, mit bestimmten Paradigmen:

- Datenbanken, Computergrafik, künstliche Intelligenz, Web-Programmierung, parallele/nebenläufige Programmierung

Organisation

- Vorlesung
- Hausaufgaben (ergibt Prüfungszulassung) mit diesen Aufgaben-Formen:
 - individuell online (autotool)
 - in Gruppen (je 3 Personen) Bearbeitung im Forum, Präsentation und Bewertung in ÜbungKoordination: Opal-Kurs (Wiki, Forum)
- Klausur: 120 min, ohne Hilfsmittel

Literatur

- <http://www.imn.htwk-leipzig.de/~waldmann/edu/ws20/pps/folien/>
- Robert W. Sebesta: Concepts of Programming Languages, Addison-Wesley 2004, ...

Zum Vergleich/als Hintergrund:

- Abelson, Sussman, Sussman: Structure and Interpretation of Computer Programs, MIT Press 1984 <http://mitpress.mit.edu/sicp/>
- Turbak, Gifford: Design Concepts of Programming Languages, MIT Press 2008 <https://cs.wellesley.edu/~fturbak/>

Inhalt

(nach Sebesta: Concepts of Programming Languages)

- Methoden: (3) Beschreibung von Syntax und Semantik
- Konzepte:
 - (5) Namen, Bindungen, Sichtbarkeiten
 - (6) Typen von Daten, Typen von Bezeichnern
 - (7) Ausdrücke und Zuweisungen, (8) Anweisungen und Ablaufsteuerung, (9) Unterprogramme
- Paradigmen:
 - (12) Objektorientierung ((11) Abstrakte Datentypen)
 - (15) Funktionale Programmierung

Haus-Aufgaben

0. Lesen Sie E. W. Dijkstra: *On the foolishness of natural language programming*
<https://www.cs.utexas.edu/users/EWD/transcriptions/EWD06xx/EWD667.html>

und beantworten Sie

- womit wird “einfaches Programmieren” fälschlicherweise gleichgesetzt?
- welche wesentliche Verbesserung brachten höhere Programmiersprachen, welche Eigenschaft der Maschinensprachen haben sie trotzdem noch?
- warum sollte eine Schnittstelle *narrow* sein?
- welche formalen Notationen von Vieta, Descartes, Leibniz, Boole sind gemeint? (jeweils: Wissenschaftsbereich, (heutige) Bezeichnung der Notation, Beispiele)
- warum können Schüler heute das lernen, wozu früher nur Genies in der Lage waren?
- Übersetzen Sie den Satz “the naturalness of ... obvious”.

Geben Sie dazu jeweils an:

- die Meinung des Autors, belegt durch konkrete Textstelle und zunächst wörtliche, dann sinngemäße Übersetzung

- Beispiele aus Ihrer Erfahrung

1. zu Skriptsprachen: finde die Anzahl der "*.java"-Dateien unter \$HOME/workspace, die den Bezeichner String enthalten (oder eine ähnliche Anwendung) (Benutze eine Pipe aus drei Unix-Kommandos.)

Lösungen:

```
find workspace/ -name "*.java" | xargs grep -l String | wc -l
find workspace/ -name "*.java" -exec grep -l String {} \; | wc -l
```

Das dient als Wiederholung zur Benutzung von Unix (GNU/Linux): führen Sie vor:

- eine Shell öffnen
- in der Manpage von find die Beschreibung von -exec anzeigen. Finden Sie (mit geeignetem Shell-Kommandos) den Quelltext dieser Manpage, zeigen diesen an. (Wie benutzt man man? so: man man.)
- was bedeutet der senkrechte Strich? in welcher Manpage steht das? in welcher Vorlesung war das dran?
- erklären Sie <https://xkcd.com/378/>, führen Sie die vier genannten Editoren vor, in dem Sie jeweils eine einzeilige Textdatei erzeugen.

Bei Vorführung (dann mit Screen-Sharing)

- schwarze Schrift auf weißem Grund
- große Schrift

2. funktionales Programmieren in Haskell (<http://www.haskell.org/>)

```
ghci
:set +t
length $ takeWhile (== '0') $ reverse $ show $ product [ 1 .. 100 ]
```

- zeigen Sie (den Studenten, die das noch nicht gesehen haben), wo die Software (hier ghc) im Pool installiert ist, und wie man sie benutzt und die Benutzung vereinfacht (PATH)
- Werten Sie den angegebenen Ausdruck aus sowie alle Teilausdrück ([1..100], product [1..100], usw.

- den Typ von `reverse` durch `ghci` anzeigen lassen
- nach diesem Typ in <https://hoogle.haskell.org/> suchen. (Einschränken auf `package:base`) Die anderen (drei) Funktionen dieses Typs aufrufen.
- eine davon erzeugt unendliche Listen, wie werden die im Speicher repräsentiert, wie kann man sie benutzen? (Am Beispiel zeigen.)

3. PostScript

```
42 42 scale 7 9 translate .07 setlinewidth .5 setgray/c{arc clip fill
setgray}def 1 0 0 42 1 0 c 0 1 1{0 3 3 90 270 arc 0 0 6 0 -3 3 90 270
arcn 270 90 c -2 2 4{-6 moveto 0 12 rlineto}for -5 2 5{-3 exch moveto
9 0 rlineto}for stroke 0 0 3 1 1 0 c 180 rotate initclip}for showpage
```

In eine Text-Datei `what.ps` schreiben (vgl. Aufgabe 1) ansehen mit `gv what.ps` (im Menu: State → watch file).

Mit Editor Quelltext ändern, Wirkung betrachten.

- Ändern Sie die Strich-Stärke!
- wie funktioniert die Steuerung einer Zählschleife?
- warum ist PostScript: imperativ? strukturiert? prozedural?
- führen Sie wenigstens ein weiteres ähnliches PostScript-Programm vor (kurzer Text, aber nichttriviale Rechnung). Quelle angeben, Programmtext erklären!
- nennen Sie einige Aspekte von PS, die in PDF übernommen wurden (Beantworten Sie anhand der Original-Dokumentation.)
- Warum sollte man niemals “online und ganz umsonst PS to PDF converter” benutzen?

2 Syntax von Programmiersprachen

Programme als Bäume

- ein Programmtext repräsentiert eine Hierarchie (einen Baum) von Teilprogrammen
- Die Semantik des Programmes wird durch Induktion über diesen Baum definiert.

- In den Blättern des Baums stehen Token,
- jedes Token hat einen Typ und einen Inhalt (eine Zeichenkette).
- dieses Prinzip kommt aus der Mathematik (arithmetische Ausdrücke, logische Formeln — sind Bäume)

Token-Typen

- reservierte Wörter (if, while, class, ...)
- Bezeichner (foo, bar, ...)
- Literale für ganze Zahlen, Gleitkommazahlen, Strings, Zeichen, ...
- Trenn- und Schlußzeichen (Komma, Semikolon)
- Klammern (runde: paren(these)s, eckige: brackets, geschweifte: braces, spitze: angle brackets)
- Operatoren (=, +, &&, ...)
- Leerzeichen, Kommentare (whitespace)

alle Token eines Typs bilden eine *formale Sprache*.

Formale Sprachen

- ein *Alphabet* ist eine Menge von Zeichen,
- ein *Wort* ist eine Folge von Zeichen,
- eine *formale Sprache* ist eine Menge von Wörtern.

Beispiele:

- Alphabet $\Sigma = \{a, b\}$,
- Wort $w = ababaaab$,
- Sprache $L =$ Menge aller Wörter über Σ gerader Länge.
- Sprache (Menge) aller Gleitkomma-Literale in \mathbb{C} .

Spezifikation formaler Sprachen

man kann eine formale Sprache beschreiben:

- *algebraisch* (Sprach-Operationen)
Bsp: reguläre Ausdrücke
- *generativ* (Grammatik), Bsp: kontextfreie Grammatik,
- durch *Akzeptanz* (Automat), Bsp: Kellerautomat,
- *logisch* (Eigenschaften), $\left\{ w \mid \forall p, r : \left(\begin{array}{l} (p < r \wedge w[p] = a \wedge w[r] = c) \\ \Rightarrow \exists q : (p < q \wedge q < r \wedge w[q] = b) \end{array} \right) \right\}$

Sprach-Operationen

Aus Sprachen L_1, L_2 konstruiere:

- Mengenoperationen
 - Vereinigung $L_1 \cup L_2$,
 - Durchschnitt $L_1 \cap L_2$, Differenz $L_1 \setminus L_2$;
- Verkettung $L_1 \cdot L_2 = \{w_1 \cdot w_2 \mid w_1 \in L_1, w_2 \in L_2\}$
- Stern (iterierte Verkettung) $L_1^* = \bigcup_{k \geq 0} L_1^k$

Def: Sprache *regulär* : \iff kann durch diese Operationen aus endlichen Sprachen konstruiert werden.

Satz: Durchschnitt und Differenz braucht man dabei nicht.

Reguläre Sprachen/Ausdrücke

Die Menge $E(\Sigma)$ der *regulären Ausdrücke* über einem Alphabet (Buchstabenmenge) Σ ist die kleinste Menge E , für die gilt:

- für jeden Buchstaben $x \in \Sigma : x \in E$
(autotool: Ziffern oder Kleinbuchstaben)
- das leere Wort $\epsilon \in E$ (autotool: `eps`)
- die leere Menge $\emptyset \in E$ (autotool: `empty`)
- wenn $A, B \in E$, dann
 - (Verkettung) $A \cdot B \in E$ (autotool: `*` oder weglassen)

- (Vereinigung) $A + B \in E$ (autotool: +)
- (Stern, Hülle) $A^* \in E$ (autotool: ^*)

Jeder solche Ausdruck beschreibt eine *reguläre Sprache*.

Beispiele/Aufgaben zu regulären Ausdrücken

Wir fixieren das Alphabet $\Sigma = \{a, b\}$.

- alle Wörter, die mit a beginnen und mit b enden: $a\Sigma^*b$.
- alle Wörter, die wenigstens drei a enthalten $\Sigma^*a\Sigma^*a\Sigma^*a\Sigma^*$
- alle Wörter mit gerade vielen a und beliebig vielen b ?
- Alle Wörter, die ein aa oder ein bb enthalten: $\Sigma^*(aa \cup bb)\Sigma^*$
- (Wie lautet das Komplement dieser Sprache?)

Erweiterte reguläre Ausdrücke

1. zusätzliche Operatoren (Durchschnitt, Differenz, Potenz), die trotzdem nur reguläre Sprachen erzeugen

Beispiel: $\Sigma^* \setminus (\Sigma^*ab\Sigma^*)^2$

ähnlich in Konfiguration der autotool-Aufgaben

2. zusätzliche nicht-reguläre Operatoren

Beispiel: exakte Wiederholungen $L^{\boxed{k}} := \{w^k \mid w \in L\}$

Bsp.: $(ab^*)^{\boxed{2}} = \{aa, abab, abbabb, ab^3ab^3, \dots\} \notin$

3. Markierung von Teilwörtern, definiert (evtl. nicht-reguläre) Menge von Wörtern mit Positionen darin

Implementierung regulärer Ausdrücke

- die richtige Methode ist Kompilation des RE in einen endlichen Automaten
Ken Thompson: *Regular expression search algorithm*, Communications of the ACM 11(6) (June 1968)
- wenn nicht-reguläre Sprachen entstehen können (durch erweiterte RE), ist keine effiziente Verarbeitung (mit endlichen Automaten) möglich.
- auch reguläre Operatoren werden gern schlecht implementiert.
Russ Cox: *Regular Expression Matching Can Be Simple And Fast (but is slow in Java, Perl, PHP, Python, Ruby, ...)*, 2007 <https://swtch.com/~rsc/regexp/regexpl.html>

Bemerkung zu Reg. Ausdr.

Wie beweist man $w \in L(X)$?

(Wort w gehört zur Sprache eines regulären Ausdrucks X)

- wenn $X = X_1 + X_2$:
beweise $w \in L(X_1)$ oder beweise $w \in L(X_2)$
- wenn $X = X_1 \cdot X_2$:
zerlege $w = w_1 \cdot w_2$ und beweise $w_1 \in L(X_1)$ und beweise $w_2 \in L(X_2)$.
- wenn $X = X_1^*$:
wähle einen Exponenten $k \in \mathbb{N}$ und beweise $w \in L(X_1^k)$ (nach vorigem Schema)

Beispiel: $w = abba$, $X = (ab^*)^*$.

$$w = abb \cdot a = ab^2 \cdot ab^0 \in ab^* \cdot ab^* \subseteq (ab^*)^2 \subseteq (ab^*)^*.$$

Übungen zu Lexik (Testfragen)

(ohne Wertung, zur Wiederholung und Unterhaltung)

- was ist jeweils Eingabe und Ausgabe für: lexikalische Analyse, syntaktische Analyse?
- warum werden reguläre Ausdrücke zur Beschreibung von Tokenmengen verwendet? (was wäre die einfachste Alternative? für welche Tokentypen funktioniert diese?)
- $(\Sigma^*, \cdot, \epsilon)$ ist Monoid, aber keine Gruppe

- $(\text{Pow}(\Sigma^*), \cup, \cdot, \dots, \dots)$ ist Halbring (ergänzen Sie die neutralen Elemente)
- In jedem Monoid: Damit $a^{b+c} = a^b \cdot a^c$ immer gilt, muß man a^0 wie definieren?
- Suchen und diskutieren Sie „Wadler’s law (of language design)“.
Am Entwurf welcher Programmiersprachen war der Autor beteiligt?

Aufgaben zu regulären Ausdrücken: autotool. Das ist Wiederholung aus VL Theoretische Informatik—Automaten und Formale Sprachen. Fragen dazu notfalls im Opal-Kurs/Forum.

Übungen zu Lexik (Hausaufgaben)

1. Für jedes Monoid $M = (D, \cdot, 1)$ definieren wir die Teilbarkeits-Relation $u \mid w := \exists v : u \cdot v = w$

Geben Sie Beispiele $u \mid w, \neg(u \mid w)$ an in den Monoiden

- $(\mathbb{N}, +, 0)$
- $(\mathbb{Z}, +, 0)$
- $(\mathbb{N}, \cdot, 1)$
- $(\{a, b\}^*, \cdot, \epsilon)$
- $(2^{\mathbb{N}}, \cup, \emptyset)$

Zeigen Sie (nicht für ein spezielles Monoid, sondern allgemein): die Relation \mid ist reflexiv und transitiv.

Ist sie antisymmetrisch? (Beweis oder Gegenbeispiel.)

NB: Beziehung zur Softwaretechnik:

- „Monoid“ ist die Schnittstelle (API, abstrakter Datentyp),
- $(\mathbb{N}, 0, +)$ ist eine Implementierung (konkreter Datentyp).
- „allgemein zeigen“ bedeutet: nur die in den Axiomen des ADT (API-Beschreibung) genannten Eigenschaften benutzen

2. Zeichnen Sie jeweils das Hasse-Diagramm dieser Teilbarkeitsrelation

- für $(\mathbb{N}, +, 0)$, eingeschränkt auf $\{0, 1, \dots, 4\}$
- für $(\mathbb{N}, \cdot, 1)$, eingeschränkt auf $\{0, 1, \dots, 10\}$

- für $(2^{\{p,q,r\}}, \cup, \emptyset)$
- für $(\{a, b\}^*, \cdot, \epsilon)$ auf $\{a, b\}^{\leq 2}$

Geben Sie eine Halbordnung auf $\{0, 1, 2\}^2$ an, deren Hasse-Diagramm ein auf der Spitze stehendes Quadratnetz ist.

Diese Halbordnung soll *intensional* angegeben werden (durch eine Formel), nicht *extensional* (durch Aufzählen aller Elemente).

3. Führen Sie vor (auf Rechner im Pool Z430, vorher von außen einloggen und probieren)

Editieren, Kompilieren, Ausführen eines kurzen (maximal 3 Zeilen) Pascal-Programms

Der Compiler `fpc` (<https://www.freepascal.org/>) ist installiert (`/usr/local/waldman`)

(Zweck dieser Teilaufgabe ist nicht, daß Sie Pascal lernen, sondern der Benutzung von `ssh`, evtl. `tmux`, Kommandozeile (PATH), Text-Editor wiederholen)

Zu regulären Ausdrücke für Tokenklassen in der Standard-Pascal-Definition <http://www.standardpascal.org/iso7185.html>

Welche Notation wird für unsere Operatoren `+` und Stern benutzt? Was bedeuten die eckigen Klammern?

In Ihrem Beispiel-Programm: erproben Sie mehrere (korrekte und fehlerhafte) Varianten für Gleitkomma-Literale. Vergleichen Sie Spezifikation (geben Sie den passenden Abschnitt der Sprachdefinition an) und Verhalten des Compilers.

Dieser Compiler (`fpc`) ist in Pascal geschrieben. Was bedeutet das für: Installation des Compilers, Entwicklung des Compilers?

4. Führen Sie vor (wie und warum: siehe Bemerkungen vorige Aufgabe): Editieren, Kompilieren (`javac`), Ausführen (`java`) eines kurzen (maximal 3 Zeilen) Java-Programms.

Suchen und buchmarken Sie die *Java Language Specification* (Primärquelle in der aktuellen Version) Beantworten Sie *damit* (und nicht mit Hausaufgabenwebseiten und anderen Sekundärquellen):

gehören in Java

- `null`
- Namen für Elemente von Aufzählungstypen

zum Tokentyp Literal, reserviertes Wort (Schlüsselwort), Bezeichner (oder evtl. anderen)?

Wo stehen die Token-Definitionen im javac-Compiler? <https://hg.openjdk.java.net/jdk/jdk15/file/>

In Ihrem Beispiel-Programm: erproben Sie verschiedene Varianten von Ganzzahl-Literalen (siehe vorige Aufgabe)

5. (erst in der darauffolgenden Woche) Gelten die Aussagen von Cox (2007) („but it’s slow in...“) jetzt immer noch? Überprüfen Sie das praktisch.

Wort-Ersetzungs-Systeme

Berechnungs-Modell (Markov-Algorithmen)

- Zustand (Speicherinhalt): Zeichenfolge (Wort)
- Schritt: Ersetzung eines Teilwortes

Syntax: Programm ist Regelmenge $R \subseteq \Sigma^* \times \Sigma^*$

Semantik: die 1-Schritt-Ableitungsrelation \rightarrow_R

$u \rightarrow_R v \iff \exists x, z \in \Sigma^*, (l, r) \in R : u = x \cdot l \cdot z \wedge x \cdot r \cdot z = v.$

Beispiele:

- Bubble-Sort: $\{ba \rightarrow ab, ca \rightarrow ac, cb \rightarrow bc\}$
- Potenzieren: $ab \rightarrow bba$ (Details: Übung)
- gibt es unendlich lange Ableitungen für: $R_1 = \{1000 \rightarrow 0001110\}, R_2 = \{aabb \rightarrow bbbaaa\}$?

Grammatiken

Grammatik G besteht aus:

- Terminal-Alphabet Σ
(üblich: Kleinbuchst., Ziffern)
- Variablen-Alphabet V
(üblich: Großbuchstaben)
- Startsymbol $S \in V$
- Regelmenge
(Wort-Ersetzungs-System)
 $R \subseteq (\Sigma \cup V)^* \times (\Sigma \cup V)^*$

Grammatik

```
{ terminale
  = mkSet "abc"
, variablen
  = mkSet "SA"
, start = 'S'
, regeln = mkSet
  [ ("S", "abc")
  , ("ab", "aabbA")
  , ("Ab", "bA")
  , ("Ac", "cc")
  ]
}
```

von G erzeugte Sprache: $L(G) = \{w \mid S \rightarrow_R^* w \wedge w \in \Sigma^*\}$.

Formale Sprachen: Chomsky-Hierarchie

- (Typ 0) aufzählbare Sprachen (beliebige Grammatiken, Turingmaschinen)
- (Typ 1) kontextsensitive Sprachen (monotone Grammatiken, linear beschränkte Automaten)
- (Typ 2) kontextfreie Sprachen (kontextfreie Grammatiken, Kellerautomaten)
- (Typ 3) reguläre Sprachen (rechtslineare Grammatiken, reguläre Ausdrücke, endliche Automaten)

Tokenklassen sind meist reguläre Sprachen.

Syntax von Programmiersprachen meist kontextfrei.

Zusatzbedingungen (Bsp: Benutzung von Bezeichnern nur nach Deklaration) meist Teil der statischen Semantik

Typ-3-Grammatiken

(= rechtslineare Grammatiken)

jede Regel hat die Form

- Variable \rightarrow Terminal Variable
- Variable \rightarrow Terminal
- Variable $\rightarrow \epsilon$

(vgl. lineares Gleichungssystem)

Beispiele

- $G_1 = (\{a, b\}, \{S, T\}, S, \{S \rightarrow \epsilon, S \rightarrow aT, T \rightarrow bS\})$
- $G_2 = (\{a, b\}, \{S, T\}, S, \{S \rightarrow \epsilon, S \rightarrow aS, S \rightarrow bT, T \rightarrow aT, T \rightarrow bS\})$

Sätze über reguläre Sprachen

Für jede Sprache L sind die folgenden Aussagen äquivalent:

- es gibt einen regulären Ausdruck X mit $L = L(X)$,
- es gibt eine Typ-3-Grammatik G mit $L = L(G)$,
- es gibt einen endlichen Automaten A mit $L = L(A)$.

Beweispläne:

- Grammatik \leftrightarrow Automat (Variable = Zustand)
- Ausdruck \rightarrow Automat (Teilbaum = Zustand)
- Automat \rightarrow Ausdruck (dynamische Programmierung)

$L_A(p, q, r) =$ alle Pfade von p nach r über Zustände $\leq q$.

Kontextfreie Sprachen

Def (Wdhlg): G ist kontextfrei (Typ-2), falls $\forall (l, r) \in R(G) : l \in V^1$
geeignet zur Beschreibung von Sprachen mit hierarchischer Struktur.

Anweisung \rightarrow Bezeichner = Ausdruck

| if Ausdruck then Anweisung else Anweisung

Ausdruck \rightarrow Bezeichner | Literal

| Ausdruck Operator Ausdruck

Bsp: korrekt geklammerte Ausdrücke: $G = (\{a, b\}, \{S\}, S, \{S \rightarrow aSbS, S \rightarrow \epsilon\})$.

Bsp: Palindrome: $G = (\{a, b\}, \{S\}, S, \{S \rightarrow aSa, S \rightarrow bSb, S \rightarrow \epsilon\})$.

Bsp: alle Wörter w über $\Sigma = \{a, b\}$ mit $|w|_a = |w|_b$

Klammer-Sprachen

Abstraktion von vollständig geklammerten Ausdrücke mit zweistelligen Operatoren

$(4 * (5+6) - (7+8)) \Rightarrow (() ()) \Rightarrow aababb$

Höhendifferenz: $h : \{a, b\}^* \rightarrow \mathbb{Z} : w \mapsto |w|_a - |w|_b$

Präfix-Relation: $u \leq w : \iff \exists v : u \cdot v = w$

Dyck-Sprache: $D = \{w \mid h(w) = 0 \wedge \forall u \leq w : h(u) \geq 0\}$

CF-Grammatik: $G = (\{a, b\}, \{S\}, S, \{S \rightarrow \epsilon, S \rightarrow aSbS\})$

Satz: $L(G) = D$. Beweis (Plan):

$L(G) \subseteq D$ Induktion über Länge der Ableitung

$D \subseteq L(G)$ Induktion über Wortlänge

(erweiterte) Backus-Naur-Form

- Noam Chomsky: Struktur natürlicher Sprachen (1956)
- John Backus, Peter Naur: Definition der Syntax von Algol (1958)

Backus-Naur-Form (BNF) \approx kontextfreie Grammatik

```
<assignment> -> <variable> = <expression>
<number> -> <digit> <number> | <digit>
```

Erweiterte BNF

- Wiederholungen (Stern, Plus) $\langle \text{digit} \rangle^+$
- Auslassungen
if $\langle \text{expr} \rangle$ then $\langle \text{stmt} \rangle$ [else $\langle \text{stmt} \rangle$]

kann in BNF übersetzt werden

Ableitungsbäume für CF-Sprachen

Def: ein geordneter Baum T mit Markierung $m : T \rightarrow \Sigma \cup \{\epsilon\} \cup V$ ist Ableitungsbaum für eine CF-Grammatik G , wenn:

- für jeden inneren Knoten k von T gilt $m(k) \in V$
- für jedes Blatt b von T gilt $m(b) \in \Sigma \cup \{\epsilon\}$
- für die Wurzel w von T gilt $m(w) = S(G)$ (Startsymbol)
- für jeden inneren Knoten k von T mit Kindern k_1, k_2, \dots, k_n gilt $(m(k), m(k_1)m(k_2) \dots m(k_n)) \in R(G)$ (d. h. jedes $m(k_i) \in V \cup \Sigma$)
- für jeden inneren Knoten k von T mit einzigem Kind $k_1 = \epsilon$ gilt $(m(k), \epsilon) \in R(G)$.

Ableitungsbäume (II)

- Def: der *Rand* eines geordneten, markierten Baumes (T, m) ist die Folge aller Blatt-Markierungen (von links nach rechts).
- Beachte: die Blatt-Markierungen sind $\in \{\epsilon\} \cup \Sigma$, d. h. Terminalwörter der Länge 0 oder 1.
- Für Blätter: $\text{rand}(b) = m(b)$,
- für innere Knoten: $\text{rand}(k) = \text{rand}(k_1) \text{rand}(k_2) \dots \text{rand}(k_n)$
- Satz: $w \in L(G) \iff$ existiert Ableitungsbaum (T, m) für G mit $\text{rand}(T, m) = w$.

Eindeutigkeit

- Def: G heißt *eindeutig* : $\forall w \in L(G) \exists$ *genau ein* Ableitungsbaum (T, m) für G mit $\text{rand}(T, m) = w$.

Bsp: $(\{a, b\}, \{S\}, S, \{S \rightarrow aSb | SS | \epsilon\})$ ist mehrdeutig.

(beachte: mehrere Ableitungen $S \rightarrow_R^* w$ sind erlaubt und wg. Kontextfreiheit auch gar nicht zu vermeiden.)

- Die naheliegende Grammatik für arith. Ausdr.

$\text{expr} \rightarrow \text{number} \mid \text{expr} + \text{expr} \mid \text{expr} * \text{expr}$

ist mehrdeutig (aus *zwei* Gründen!) — Auswege:

- Transformation zu eindeutiger Grammatik (benutzt zusätzliche Variablen)
- Operator-Assoziativitäten und -Präzedenzen

Assoziativität

- (Wdhlg.) Definition: Operation ist *assoziativ*
- für nicht assoziativen Operator \odot muß man festlegen, was $x \odot y \odot z$ bedeuten soll:

$$(3 + 2) + 4 \stackrel{?}{=} 3 + 2 + 4 \stackrel{?}{=} 3 + (2 + 4)$$

$$(3 - 2) - 4 \stackrel{?}{=} 3 - 2 - 4 \stackrel{?}{=} 3 - (2 - 4)$$

$$(3 * * 2) * * 4 \stackrel{?}{=} 3 * * 2 * * 4 \stackrel{?}{=} 3 * *(2 * * 4)$$

- ... und dann die Grammatik entsprechend einrichten
(d.h., eine äquivalente eindeutige Grammatik konstruieren, deren Ableitungsbäume die gewünschte Struktur haben)

Assoziativität (II)

- $X1 - X2 + X3$ auffassen als $(X1 - X2) + X3$
- Grammatik-Regeln

Ausdruck \rightarrow Zahl | Ausdruck + Ausdruck
| Ausdruck - Ausdruck

- ersetzen durch

Ausdruck \rightarrow Summe
Summe \rightarrow Summand | Summe + Summand
| Summe - Summand
Summand \rightarrow Zahl

Präzedenzen

- Beispiel

$$(3 + 2) * 4 \stackrel{?}{=} 3 + 2 * 4 \stackrel{?}{=} 3 + (2 * 4)$$

- Grammatik-Regel

summand \rightarrow zahl

- erweitern zu

summand \rightarrow zahl | produkt
produkt \rightarrow ...

(Assoziativität beachten)

Zusammenfassung Operator/Grammatik

Ziele:

- Klammern einsparen
- trotzdem eindeutig bestimmter Syntaxbaum

Festlegung:

- Assoziativität: bei Kombination eines Operators mit sich
- Präzedenz: bei Kombination verschiedener Operatoren

Realisierung in CFG:

- Links/Rechts-Assoziativität \Rightarrow Links/Rechts-Rekursion
- verschiedene Präzedenzen \Rightarrow verschiedene Variablen

Hausaufgaben (Rewriting, Typ-0-Grammatiken)

- Für das $R = \{ab \rightarrow baa\}$ über $\Sigma = \{a, b\}$:

was ist die R -Normalform von

- a^3b , allgemein $a^k b$,
- ab^3 , allgemein ab^k ,

formulieren Sie die allgemeinen Aussagen exakt, überprüfen Sie für $k = 3$, beweisen Sie durch vollständige Induktion.

Hausaufgaben (CF-Grammatiken)

1. zur Geschichte:

- Definitionen von CFG und Ableitungsbaum in Noam Chomsky: *Three Models for the Description of Language*, 1956 <https://chomsky.info/articles/>. Wie heißt CFG dort?
- vergleichen Sie die Syntax-Definitionen von Fortran (John Backus 1956) und Algol (Peter Naur 1960),
Quellen siehe <http://web.eah-jena.de/~kleine/history/>

2. für die offizielle Java-Grammatik (nach JLS in aktueller Version)

- es werden tatsächlich zwei Grammatiken benutzt (lexikalische, syntaktische), zeigen Sie deren Zusammenwirken an einem einfachen Beispiel (eine Ableitung, bei der in jeder Grammatik nur wenige Regeln benutzt werden)
- bestimmen Sie den Ableitungsbaum (bzgl. der syntaktischen Grammatik) für das übliche `hello world`-Programm,
- Beispiele in `jshell` vorführen. Wie lautet die Grammatik für die dort erlaubten Eingaben? Falls es sie gibt - sonst eine andere primäre Dokumentations dafür.

3. bzgl. der eindeutigen Grammatik für arithmetische Ausdrücke, die in der VL (Skript) entwickelt wurde:

- Ableitungsbaum für $1 * 2 - 3 * 4$

- Grammatik erweitern für geklammerte Ausdrücke, Eindeutigkeit begründen, Ableitungsbaum für $1 * (2 - 3) * 4$ angeben
- welche Variable der offiziellen Java-Grammatik erzeugt arithmetische Ausdrücke? Ableitungsbaum für $1 * (2 - 3) * 4$ von dieser Variablen aus angeben.

4. (autotool) diese oder ähnliche Aufgaben:

- CF-Grammatik für $\{w \mid w \in \{a, b\}^*, |w|_a = |w|_b\}$
- CF-Grammatik für $\{w \mid w \in \{a, b\}^*, 2 \cdot |w|_a = |w|_b\}$
- jeweils auch als eindeutige CFG

3 Semantik von Programmiersprachen

Statische und dynamische Semantik

Semantik = Bedeutung

- statisch (kann zur Übersetzungszeit geprüft werden)

Beispiele (in C, Java, ...)

- Typ-Korrektheit von Ausdrücken,
- Bedeutung (Bindung) von Bezeichnern

- dynamisch (beschreibt Ausführung des Programms)

operational, axiomatisch, denotational

Hilfsmittel: Attributgrammatiken

Bsp statische/dynamische Semantik

Benutzung eines undeklarierten Namens:

- Java: verhindert durch statische Semantik-Prüfung (Programm wird nicht übersetzt, nicht ausgeführt)

```
{ System.out.print("foo"); System.out.println(x);
| Error:
| cannot find symbol
| symbol:   variable x
```

- ECMA-Script (Javascript): dynamische Semantik ist „Exception wird ausgelöst“ (Programm wird ausgeführt)

```
> {console.log("foo"); console.log(x);}
foo
Thrown:
ReferenceError: x is not defined
```

Attributgrammatiken (I)

- Attribut: Annotation an Knoten des Syntaxbaums.
 $A : \text{Knotenmenge} \rightarrow \text{Attributwerte}$ (Bsp: \mathbb{Z})
- Attributgrammatik besteht aus:
 - kontextfreier Grammatik G (Bsp: $\{S \rightarrow e \mid mSS\}$)
 - für jeden Knotentyp (Terminal + Regel)
 eine Menge (Relation) von erlaubten Attribut-Tupeln $(A(X_0), A(X_1), \dots, A(X_n))$
 für Knoten X_0 mit Kindern $[X_1, \dots, X_n]$

$S \rightarrow mSS, A(X_0) + A(X_3) = A(X_2); S \rightarrow e, A(X_0) = A(X_1);$ Terminale:
 $A(e) = 1, A(m) = 0$

Attributgrammatiken (II)

ein Ableitungsbaum mit Annotationen ist *korrekt bezüglich einer Attributgrammatik*, wenn

- er zur zugrundeliegenden CF-Grammatik paßt
- in jedem Knoten das Attribut-Tupel (von Knoten und Kindern) zur erlaubten Tupelmenge gehört

Plan:

- Baum beschreibt Syntax, Attribute beschreiben Semantik

Ursprung: Donald Knuth: Semantics of Context-Free Languages, (Math. Systems Theory 2, 1968)

technische Schwierigkeit: Attributwerte effizient bestimmen. (beachte: (zirkuläre) Abhängigkeiten)

Donald E. Knuth

- The Art Of Computer Programming (1968, ...) (Band 3: Sortieren und Suchen)
- T_EX, Metafont, Literate Programming (1983, ...) (Leslie Lamport: L^AT_EX)
- Attribut-Grammatiken (1968)
- Anwendung der Landau-Notation ($O(f)$, Analysis) und Erweiterung (Ω , Θ) für asymptotische Komplexität
- ...

<https://www-cs-faculty.stanford.edu/~uno/>

Arten von Attributen

- synthetisiert:
hängt nur von Attributwerten in Kindknoten ab
Bsp: Typ von Ausdrücken, Wert von Ausdrücken
- ererbt (inherited)
hängt nur von Attributwerten in Elternknoten und (linken) Geschwisterknoten ab
Bsp: deklarierte Typen für Namen

Wenn Abhängigkeiten bekannt sind, kann man Attributwerte durch Werkzeuge bestimmen lassen.

(Bransen et al.: *Linearly Ordered Attribute Grammar Scheduling ...*, TACAS 2015
https://doi.org/10.1007/978-3-662-46681-0_24)

Attributgrammatiken–Beispiele

- Auswertung arithmetischer Ausdrücke (dynamisch)
- Bestimmung des abstrakten Syntaxbaumes
- Typprüfung (statisch)
- Kompilation (für Kellermaschine) (statisch)

Konkrete und abstrakte Syntax

- konkreter Syntaxbaum = der Ableitungsbaum
- abstrakter Syntaxbaum = wesentliche Teile des konkreten Baumes

unwesentlich sind z. B. die Knoten, die zu Hilfsvariablen der Grammatik gehören.
abstrakter Syntaxbaum kann als synthetisiertes Attribut konstruiert werden.

```
E -> E + P ; E.abs = new Plus (E.abs, P.abs)
E -> P      ; E.abs = P.abs
```

Regeln zur Typprüfung

... bei geschachtelten Funktionsaufrufen

- Funktion f hat Typ $A \rightarrow B$
- Ausdruck X hat Typ A
- dann hat Ausdruck $f(X)$ den Typ B

Beispiel

```
class C {
  static class A {} static class B {}
  static B f (A y) { .. }
  static A g (B x) { .. }
  ..
  .. C.g (C.f (new C.A())) .. }
```

Bsp. Operationale Semantik: Keller

- Kellerspeicher
 - Zustand ist Zahlenfolge $s \in \mathbb{Z}^*$, Empty = []
 - Operationen:
 - * Push(x), Semantik: $[s_1, \dots, s_n] \rightarrow [x, s_1, \dots, s_n]$
 - * $y :=$ Pop(), Semantik: $[y, s_1, \dots, s_n] \rightarrow [s_1, \dots, s_n]$

- Realisierung zweistelliger Verknüpfungen: Argumente vom Keller holen, Resultat auf Keller schreiben, z.B.

Plus $\equiv \{a := \text{Pop}(); b := \text{Pop}(); \text{Push}(a + b)\}$

- benutzt in Prog.-Spr. Forth (1970), PostScript (1982), JVM (Java Virtual Machine, 1994), Bsp: 6.5 iadd

Kompilation für Kellermaschine

- Spezifikation:
 - Eingabe: Java-Ausdruck A , Bsp. $3 * x + 1$
 - Ausgabe: JVM-Programm P , Bsp:


```
push 3; push x; imul; push 1; iadd;
```
 - Zusammenhang: $\square \xrightarrow{P} [\text{Wert}(A)]$
 - dann gilt auch $\forall k \in \mathbb{Z}^* : k \xrightarrow{P} ([\text{Wert}(A)] \circ k)$
- Realisierung (Kompilation):
 - Code für Konstante/Variable c : `push c;`
 - Code für Ausdruck $x \circ y$: `code(x); code(y); o;`

der so erzeugte Code ist synthetisiertes Attribut
- JVM-Programm (Bytecode) ansehen mit `javap -c`,

Attributgrammatiken mit SableCC

- SableCC: <https://sablecc.org/>
SableCC is a parser generator for building compilers, interpreters ..., strictly-typed abstract syntax trees and tree walkers
- Syntax einer Regel

```
linke-seite { -> attribut-typ }
  = { zweig-name } rechte-seite { -> attribut-wert }
```

- Quelltexte: `git clone https://gitlab.imn.htwk-leipzig.de/waldmann/pps-ws15`
Benutzung: `cd pps-ws15/rechner ; make ; make test ; make clean`
(dafür muß `sablecc` gefunden werden, siehe <http://www.imn.htwk-leipzig.de/~waldmann/etc/pool/>)
- Struktur:
 - `rechner.grammar` enthält Attributgrammatik, diese beschreibt die Konstruktion des *abstrakten Syntaxbaumes (AST)* aus dem Ableitungsbaum (konkreten Syntaxbaum)
 - `Eval.java` enthält Besucherobjekt, dieses beschreibt die Attributierung der AST-Knoten durch Zahlen
 - Hauptprogramm in `Interpreter.java`
 - bauen, testen, aufräumen: siehe `Makefile`
 - generierte Dateien in `rechner/*`

Bemerkungen (häufige/nicht offensichtliche Fehlerquellen)

- Redefinition of `... : nicht so: foo -> bar ; foo -> baz; sondern so: foo -> {eins} bar | {zwei} baz;`
Regeln mit gleicher linker Seite zusammenfassen,
die rechten Seiten durch Label (`{eins}`, `{zwei}`) unterscheiden
- `... conflict ...:`
die Grammatik ist nicht eindeutig (genauer: wird von Sablecc nicht als eindeutig erkannt)

Kommentar: in Java fehlen: algebraische Datentypen, Pattern Matching, Funktionen höherer Ordnung. Deswegen muß SableCC das simulieren — das sieht nicht schön aus. Die „richtige“ Lösung sehen Sie später im Compilerbau.

Abstrakter Syntaxbaum, Interpreter: <http://www.imn.htwk-leipzig.de/~waldmann/edu/ws11/cb/folien/main/node12.html>, Kombinator-Parser: <http://www.imn.htwk-leipzig.de/~waldmann/edu/ws11/cb/folien/main/node70.html>

Hausaufgaben

1. die angegebene Sablecc-Grammatik ergänzen: Multiplikation, Subtraktion, Klammern
2. `javap -c` vorführen für Übersetzung von arithmetischen Ausdrücken (keine Programmablaufsteuerung)
Ausgabe vergleichen mit dem Verfahren aus VL
welchen Maschinencode erzeugt der JIT-Compiler der JVM daraus? (<https://wiki.openjdk.java.net/display/HotSpot/PrintAssembly>, ist/wird im Pool installiert)
3. ggf. Effizienz der Impl. reg. Ausdr. (alte Aufgabe)

Dynamische Semantik

- operational:
beschreibt Wirkung einzelner Anweisungen durch Änderung des Speicherbelegung
- denotational:
ordnet jedem (Teil-)Programm einen Wert zu, Bsp: eine Funktion (höherer Ordnung).
Beweis von Programmeigenschaften durch Term-Umformungen
- axiomatisch (Bsp: Hoare-Kalkül):
enthält Schlußregeln, um Aussagen über Programme zu beweisen

Operationale Semantik: Sprünge

- Maschinenmodell:
Variable `PC` (program counter) enthält Adresse des nächsten auszuführenden Befehls
- Semantik von `Goto(z)` ist: $PC := z$
Semantik der Nicht-Sprungbefehle: $\dots, PC := PC + 1$
- andere Varianten der Programmablaufsteuerung können in Goto-Programme übersetzt werden
Bsp: Schleife `while (B) A` \Rightarrow `if (B) ...`
das findet bei Kompilation von Java nach JVM statt

Denotationale Semantik

Beispiele: Semantik von ... ist ...

- (nebenwirkungsfreies) *Unterprogramm* \Rightarrow Funktion von Argument nach Resultat
- *Anweisung* \Rightarrow Funktion von Speicherbelegung nach Speicherbelegung

Vorteile denotationaler Semantik:

- Bedeutung eines Programmes = mathematisches Objekt
- durch Term beschreiben, durch äquivalente Umformungen verarbeiten (equational reasoning)

Vorteil deklarativer Programmierung: Programmiersprache *ist* Beschreibungssprache (*ist* Mathematik)

Beispiele Denotationale Semantik

- jeder arithmetische Ausdruck (aus Konstanten und Operatoren)
beschreibt eine Zahl
- jeder aussagenlogische Ausdruck (aus Variablen und Operatoren)
beschreibt eine Funktion (von Variablenbelegung nach Wahrheitswert)
- jeder reguläre Ausdruck
beschreibt eine formale Sprache
- jedes rekursiv definierte Unterprogramm
beschreibt eine *partielle* Funktion

Beispiel: Denotationale Sem. von Unterprogr.

- Unterprogramme definiert durch Gleichungssysteme.
Sind diese immer lösbar? (eindeutig?)
- Diese f, t sind tatsächlich einfach darstellbar:

```
f (x) = if x > 52
      then x - 11 else f (f (x + 12))
```

```
t (x, y, z) = if x <= y then z + 1
             else t ( t (x-1, y, z)
                    , t (y-1, z, x)
                    , t (z-1, x, y) )
```

das ist aber Glück, genauer, Absicht, es sind berühmte Vorführ-Beispiele (Ü: Auto-
ren, Quellen? Hinweis: DEK)

Aufgabe Denotationale Semantik

- $g(x, y) = \text{if } x \leq 0 \text{ then } 0$
 $\text{else if } y \leq 0 \text{ then } 0$
 $\text{else } 1 + g(g(x-1, y), g(x, y-1))$
- 1. Wertetabelle durch Programm bestimmen! Programmiersprache dabei völlig egal!
 2. Selbständig!
- Wenn die Rechnung zu lange dauert: 3. verstehen, warum, 4. effizienteren Algorithmus benutzen
 welches Algorithmen-Entwurfsprinzip hilft hier?
- das ist alles Wiederholung aus der VL (*Grundlagen der*) *Programmierung* bzw. *Algorithmen und Datenstrukturen*

Axiomatische Semantik

Notation für f. Aussagen über Speicherbelegungen: *Hoare-Tripel*: $\{ V \} A \{ N \}$

- für jede Belegung s , in der Vorbedingung V gilt:
- *wenn* Anweisung A ausgeführt wird
 und Belegung t erreicht wird,
- *dann* gilt dort Nachbedingung N

Beispiel: $\{ x \geq 5 \} y := x + 3 \{ y \geq 7 \}$

Gültigkeit solcher Aussagen kann man

- beweisen (mit Hoare-Kalkül)
- prüfen (testen)

Beachte: $\{x \geq 5\} \text{ while } (\text{true}) ; \{x == 42\}$

Eiffel

Bertrand Meyer, <http://www.eiffel.com/>

```
class Stack [G]      feature
  count : INTEGER
  item  : G is require not empty do ... end
  empty : BOOLEAN is do .. end
  full  : BOOLEAN is do .. end
  put (x: G) is
    require not full do ...
    ensure not empty
      item = x
      count = old count + 1
```

Beispiel sinngemäß aus: B. Meyer: Object Oriented Software Construction, Prentice Hall 1997

Hoare-Kalkül: Überblick

zu jedem Knotentyp in abstrakten Syntaxbäumen von strukturierten imperativen Programmen ein Axiom-Schema

- elementare Anweisung:
 - Zuweisung $\{ N[x/E] \} x := E \{ N \}$
- zusammengesetzte Anweisungen:
 - wenn $\{ V \} C \{ Z \}$ und $\{ Z \} D \{ N \}$
dann $\{ V \} C; D \{ N \}$
 - wenn $\{ V \text{ und } B \} C \{ N \}$
und $\{ V \text{ und not } B \} D \{ N \}$
dann $\{ V \} \text{if } (B) \text{ then } C \text{ else } D \{ N \}$
 - wenn $\{ I \text{ and } B \} A \{ I \}$,
dann $\{ I \} \text{while } (B) \text{ do } A \{ I \text{ and not } B \}$

Anwendung des Zuweisungs-Axioms

- Axiom (-Schema): $\{ N[x/E] \} x := E \{ N \}$

- dabei bedeutet $N[x/E]$:
 der Ausdruck N , wobei jedes Vorkommen des Namens x durch den Ausdruck E ersetzt wird
 Bsp: $(y \geq 7)[y/x + 3] = (x + 3 \geq 7) = (x \geq 4)$
- Bsp: Anwendung $\{ \dots \} y := x+3 \{ y \geq 7 \}$
- Übung: welche Vorbedingung ergibt sich für
 $a := a + b; b := a - b; a := a - b;$
 und Nachbedingung $a = X \wedge b = Y?$
 Dabei auch Axiom für Nacheinanderausführung benutzen

Hoare-Kalkül: Ergänzung

- logische Umformungen (Programm A bleibt erhalten)
 - Verschärfen einer Vorbedingung (von V zu V')
 - Abschwächen einer Nachbedingung (von N zu N')

wenn $\{ V \} A \{ N \}$ und $V' \Rightarrow V$ und $N \Rightarrow N'$
 dann $\{ V' \} A \{ N' \}$
- Anwendung: beweise $\{ x < 4 \} x := 5-x \{ x > 2 \}$
 - Zuweisungs-Axiom ergibt $\{ 5-x > 2 \} x := 5-x \{ x > 2 \}$
 - äquivalent umgeformt zu $\{ x < 3 \} x := 5-x \{ x > 2 \}$
 - dann o.g. Axiom anwenden mit $V = (x < 3)$, $V' = (x < 4)$, $N = N' = (x > 2)$

Anwendung: Axiom für Verzweigung

- das Axiom:

wenn $\{ V \text{ und } B \} C \{ N \}$
 und $\{ V \text{ und not } B \} D \{ N \}$
 dann $\{ V \} \text{if } (B) \text{ then } C \text{ else } D \{ N \}$
- Anwendung: beweisen Sie
 $\{ x > 9 \}$
 $\text{if } (x > y) \text{ then } a := x - 2 \text{ else } a := y + 2$
 $\{ a > 7 \}$

Axiom für Verzweigung (Rechnung)

- wir müssen $\{x > 9 \text{ und } x > y\} \ a := x - 2 \ \{a > 7\}$
und $\{x > 9 \text{ und } x \leq y\} \ a := y + 2 \ \{a > 7\}$ zeigen,
um das Axiom-Schema anwenden zu können
- Zuweisungs-Axiom ergibt $\{x - 2 > 7\} \ a := x - 2 \ \{a > 7\}$
äquivalent umgeformt $\{x > 9\} \ a := x - 2 \ \{a > 7\}$
Axiom-Schema zum Verschärfen der Vorbedingung ($V' = (x > 9) \wedge (x > y)$, $V = (x > 9)$) ergibt erstes Teilziel
- Zuweisungs-Axiom ergibt $\{y + 2 > 7\} \ a := y + 2 \ \{a > 7\}$
äquivalent umgeformt $\{y > 5\} \ a := y + 2 \ \{a > 7\}$
Axiom-Schema zu Verschärfen der Vorbedingung ist anwendbar für $V' = (x > 9 \wedge x \leq y)$, $V = (y > 5)$ wegen $V' \Rightarrow V$, ergibt zweites Teilziel

Anwendung des Axioms für Schleifen

- wenn $\{ I \text{ and } B \} \ A \ \{ I \}$,
dann $\{ I \} \ \text{while } (B) \ \text{do } A \ \{ I \text{ and not } B \}$
- Eingabe `int p, q; // p = P und q = Q`
`int c = 0;`
`// inv: p * q + c = P * Q`
`while (q > 0) {`
 `??? := ???; q := q - 1;`
`}`
`// c = P * Q`
- Invariante muß: 1. vor der Schleife gelten, 2. im S.-Körper invariant bleiben, 3. nach der Schleife nützlich sein
- erst Spezifikation (hier: Invariante), dann Implementierung. (sonst: *cart before the horse*, EWD 1305)

Aufgaben axiomatische Semantik

1. zur Folie „Zuweisungs-Axiom“:
bestimmen Sie die Vorbedingung zu $a := a + b; \dots$, aus den Axiomen für Zuweisung und Nacheinanderausführung.

2. Ergänzen Sie das Programm, so daß die Spezifikation (das Potenzieren) erfüllt wird.

```
Eingabe: natürliche Zahlen a, b;
// a = A und b = B
int p := 1; int c := ???;
// Invariante:  $c^b * p = A^B$ 
while (b != 0) {
    if (b ist ungerade)
        then (c,p) := ...
        else (c,p) := ...
    // Z
    b := abrunden (b/2);
}
Ausgabe: p; //  $p = A^B$ 
```

- Initialisieren Sie c so, daß die Invariante gilt.
- Wieso folgt aus der Invariante bei Verlassen der Schleife die Korrektheit der Ausgabe?
- Bestimmen Sie eine geeignete Aussage Z als Vorbedingung der nachfolgenden Anweisung bezüglich der Invariante.
- Bestimmen Sie daraus die Lücken (. . .)

Literatur/Kommentar zu Verifikation

- Einführung Hoare-Kalkül: Kap. 2.11 in Gumm/Sommer: *Informatik*, Oldenbourg 2013 (online über HTWK-Bibl.).
- Dort allerdings fällt das Programm (Schleife) vom Himmel und der Beweis (Invariante) wird nachträglich geraten.
- Vergleiche dazu E. W. Dijkstra: *putting the cart (das Programmieren) before the horse (das Beweisen)*. <https://www.cs.utexas.edu/users/EWD/transcriptions/EWD13xx/EWD1305.html>.
- In welchen Semestern finden Sie die VL zum Programmieren, die VL zur Programmverifikation? Aussch.
Vergleichen Sie mit Studiengang Bauingenieurwesen, VL Baumechanik/VL Bauwirtschaft (und ähnliche)

4 Mehr zu Korrektheit und Termination

Erweiterter Euklidischer Algorithmus (Spezif.)

- Spezifikation: (Beispiel: $x = 60, y = 35, \dots$)
 - Eingabe: $x, y \in \mathbb{N}$
 - Ausgabe: $a, b \in \mathbb{Z}$ mit $g = a \cdot x + b \cdot y \wedge g|x \wedge g|y$
- Satz: $g = \text{gcd}(x, y)$.
- Def: $\text{gcd}(x, y)$ ist das Infimum (größte untere Schranke) von x und y in der Teilbarkeits-Halbordnung, d.h.,
 $\text{gcd}(x, y)|x \wedge \text{gcd}(x, y)|y \wedge \forall h : \dots$
- Beweis des Satzes:
 1. $g|x \wedge g|y$ nach Spezifikation,
 2. \dots

Erweiterter Euklid — funktional (Aufgabe)

- Ausgabe: $a, b \in \mathbb{Z}$ mit $\text{gcd}(x, y) = a \cdot x + b \cdot y$
- ```
ee :: Int -> Int -> (Int, Int)
ee x y =
 if y == 0 then (... , ...)
 else let (d,m) = divMod x y
 (p,q) = ee y m
 in (... , ...)
```
- Beweis: Induktion nach  $y$
- benutzt Eigenschaften des gcd:  
 $\text{gcd}(x, 0) = \dots, \text{gcd}(d \cdot y + m, y) = \text{gcd}(m, y)$ .

### Erweiterter Euklid — funktional (Lösung)

- Ausgabe:  $a, b \in \mathbb{Z}$  mit  $\text{gcd}(x, y) = a \cdot x + b \cdot y$

- `ee :: Int -> Int -> (Int, Int)`  
`ee x y =`  
`if y == 0 then (1, 0)`  
`else let (d, m) = divMod x y`  
`(p, q) = ee y m`  
`in (q, p - d * q)`

### Simultan-Zuweisung

- Anweisung  $(v_1, v_2) := (e_1, e_2)$  für  $v_1 \neq v_2$
- axiomatische Semantik:  
 $\{N[v_1/e_1, v_2/e_2]\} (v_1, v_2) := (e_1, e_2) \{N\}$   
verwendet links simultane Ersetzung
- Bsp:  $\{\dots\} (a, b) := (b, a) \{a = 2 \wedge b = 5\}$   
Bsp:  $\{\dots\} (x, y) := (x + y, x - y) \{x = 7 \wedge y \geq 3\}$
- realisiert in der Sprache CPL 1963
- in JS als *destructuring assignment*, ECMA 262: 12.15.5  
`[a, b] = [8, 9]; [a, b] = [b, a]`

### Erweiterter Euklid — imperative Impl.

- Ansatz: verwende  $x, y, a, b, p, q \geq 0$  mit Invariante  
 $\text{gcd}(x, y) = \text{gcd}(x_{\text{in}}, y_{\text{in}}) \wedge x = a \cdot x_{\text{in}} + b \cdot y_{\text{in}}, y = p \cdot x_{\text{in}} + q \cdot y_{\text{in}}$
- `// X = x, Y = y, x >= 0, y >= 0`  
`(a, b, p, q) := ...`  
`// Inv: gcd(x, y) = gcd(X, Y),`  
`// x = a X + b Y, y = p X + q Y`  
`while ( y > 0 ) {`  
`(x, y, a, b, p, q) := (y, x mod y, ...)`  
`}`  
`// gcd(X, Y) = a X + b Y`

## Aufgabe zum erweiterten Eukl. Alg.

### 1. Für das Programm

```
Eingabe: positive natürliche Zahlen A, B;
(a, b, c, d) := (A, B, B, A)
while (a /= b) {
 if (a > b) then (a, d) := (a-b, c+d)
 else (b, c) := (b-a, d+c)
}
Ausgabe: (a+b)/2 , (c+d)/2
```

- zeigen Sie, daß die erste Ausgabe gleich  $\text{gcd}(A, B)$  ist.  
Beweisen Sie dazu die Invarianz von  $\text{gcd}(a, b) = \text{gcd}(A, B)$ .  
Welche Eigenschaften des  $\text{gcd}$  werden benötigt?
- was ist die zweite Ausgabe?  
Geben Sie eine Vermutung an und beweisen Sie mit einer geeigneten Invariante.
- wozu ist die Bedingung „positiv“ notwendig?

## Partielle und totale Korrektheit

- Hoare-Tripel  $\{V\} A \{N\}$  beschreibt *partielle* Korrektheit:  
wenn vorher  $V$  gilt und  $A$  hält, dann gilt danach  $N$   
Bsp:  $\{\text{true}\} \text{while } (\text{true}); \{x=42\}$  ist wahr
- stärker (und nützlicher) ist *totale* Korrektheit:  
partielle Korrektheit und  $A$  hält tatsächlich (*Termination*)
- Beweis-Verfahren (für Schleifen  $\text{while } (B) \text{ do } A$ )
  - partielle Korrektheit: Invariante
  - Termination: Maßfunktion (Schrittfunktion)  
 $m$ : Speicherbelegung  $\rightarrow \mathbb{N}$  mit  $\{m = M\} A \{m < M\}$   
Bsp: eine Maßf. für  $\text{while } (x > 0) \{ \dots; x = x/2; \}$  ist  $x$
- es gilt:  $m(\text{aktueller Zustand}) \geq \text{verbleibende Anzahl der Schleifendurchläufe (Schritte)}$

## Wie findet man die Maßfunktion?

- die richtige Antwort ist (wie für die Invariante):  
das ist die falsche Frage.
- das softwaretechnisch richtige Vorgehen ist:
  1. (Entwurf) Invariante und Maßfunktion hinschreiben,
  2. (Implementierung) Schleife so ausprogrammieren, daß diese Behauptungen stimmen.
- in der Praxis wünscht man eine Teil-Automatisierung:  
maschinelles Finden von einfachen („offensichtlichen“) Maßfunktionen und Beweisen dafür
- vgl. Typisierung von Namen:
  - Deklaration (durch Programmierer),
  - Inferenz (durch Compiler) (z.B. `var` in C#, `auto` in C++)

## Maßfunktionen für funktionale Programme

- (VAR x y) (RULES  $g(Z, y) \rightarrow Z$   $g(x, Z) \rightarrow Z$   
 $g(S(x), S(y)) \rightarrow S(g(g(x, S(y)), g(S(x), y)))$  )  
„Speicherbelegung“ ist jetzt Term-Baum
- automatische Terminations-Analyse mit TTT2 <http://colo6-c703.uibk.ac.at/ttt2/> (U. Innsbruck)  

```
$ ttt2 min.trs # YES Proof: DP Processor: ...
Matrix Interpretation Processor: ...
g#(S(x), S(y)) = 2y+2 >= 2y = g#(g(x, S(y)), g(S(x), y))
```
- genau das ist mein Forschungsgebiet <https://dblp.uni-trier.de/pers/hd/w/Waldmann:Johannes>,
- Teil der VL Symbolisches Rechnen (im SS 21) <https://informatik.htwk-leipzig.de/waldmann/lehre.html>

## Totale Programmiersprachen

- eine Sprache  $P$  heißt *total*, wenn jedes (statisch korrekte)  $P$ -(Unter)Programm  $f$  für alle Argumente terminiert  
(denotationale Semantik von  $f$  ist eine totale Funktion)
- *nicht total*: beliebige Sprünge, beliebige (bedingungsgesteuerte) Schleifen, beliebige Rekursion
- *total*: Sprünge nur nach vorn, nur Zählschleifen (bei denen Zähler im Schleifenkörper nicht geändert wird), Rekursion nur durch Schema (fold, primitive Rekursion)
- Beispiel: Agda (Ulf Norell 2007, Andreas Abel, et al.) <https://wiki.portal.chalmers.se/agda/agda.php?n=ReferenceManual.Totality>

## Ressourcenverbrauch von Programmen

- für totale Programme: Schranken f. Ressourcenverbrauch  
(max. Anzahl Rechenschritte als Fkt. der Eingabegröße)  
z.B. polynomiell, quadratisch, linear  
deklarieren (manuell) und inferieren (automatisch)
- Martin Hofman, Jan Hoffman, et al.: *Resource Aware ML*, <https://www.raml.co/publications/>
- <https://accidentallyquadratic.tumblr.com/>  
Bsp: einfach verkettete Listen:  $xs ++ ys$  ist linear in  $xs$ 
  - $(.. (xs1 ++ xs2) ++ ..) ++ xsN$ , d.h., *foldl*
  - $xs1 ++ (xs2 ++ (.. ++ xsN) ..)$ , d.h., *foldr*

## Aufgaben und Literatur zu Termination

- (für das Wahlfach *Symbolisches Rechnen* einschreiben!)
- Nachum Dershowitz: *33 Examples of Termination*, 1993. <http://www.math.tau.ac.il/~nachumd/papers/printemp-print.pdf>  
Geben Sie möglichst lange Ableitungen (von kleinem Startterm) für das Term-Ersetzungs-System 12 (disjunktive Normalform) an. Überprüfen Sie für einen Schritt  $t_1 \rightarrow_R t_2$  aus einer solchen Ableitung die Eigenschaft  $t_1 > t_2$  der angegebenen Maßfunktion.

- Alan Turing (1950): *Checking a Large Routine* <http://www.turingarchive.org/browse.php/B/8>  
Wo wird Termination behandelt? Wie? — Das ist der älteste bekannte Terminationsbeweis eines Programms!

## 5 Typen

### Warum Typen?

- Typ ist Menge von Werten mit Operationen
- für jede eigene Menge von Werten (Variablen) aus dem *Anwendungsbereich* benutze eine eigenen Typ  
... und verdeutliche das im Quelltext!
- Bezeichner werden statisch typisiert,  
der Typ der Daten zur Laufzeit (an der bezeichneten Speicherstelle) stimmt dann mit dem statischen Typ des Bezeichners überein (Sicherheit)  
und muß nicht mehr getestet werden (Effizienz)

### Historische Entwicklung

- keine Typen (nur ein Typ: alles ist Maschinenwort)
- vorgegebene Typen (Fortran: Integer, Real, Arrays)
- benutzerdefinierte Typen  
(algebraische Datentypen;  
Spezialfälle: enum, struct, class)
- abstrakte Datentypen (interface)
- polymorphe Typen (z.B. `List<E>`, Arrays, Zeiger)
- (data) dependent types (z.B. in Agda, Idris)

## Überblick

- einfache (primitive) Typen
  - Zahlen, Wahrheitswerte, Zeichen
  - benutzerdefinierte Aufzählungstypen
  - Teilbereiche
- zusammengesetzte (strukturierte) Typen
  - Produkt (record)
  - Summe (union) (Spezialfall: Aufzählungen)
  - rekursive Typen (Anwendung: Listen, Bäume)
  - Potenz (Funktionen): Unterprogramme, Arrays, (Tree/Hash-)Maps  
Verweistypen (Zeiger) als Spezialfall von Arrays

## Zahlenbereiche

- Maschinenzahlen (oft im Sprachstandard festgelegt)
  - ganze Zahlen (in binärem Zweierkomplement)
  - gebrochene Zahlen (in binärer Gleitkommadarstellung)  
Goldberg 1991: *What Every Computer Scientist Should Know About Floating-Point Arithmetic* [http://docs.oracle.com/cd/E19957-01/806-3568/ncg\\_goldberg.html](http://docs.oracle.com/cd/E19957-01/806-3568/ncg_goldberg.html)
- Abstraktionen (oft in Bibliotheken, Bsp. <https://gmpilib.org/manual/>)
  - beliebig große Zahlen
  - exakte rationale Zahlen

## Aufzählungstypen

können einer Teilmenge ganzer Zahlen zugeordnet werden

- durch Sprache vorgegeben: z.B. int, char, boolean
- anwendungsspezifische (benutzerdef.) Aufzählungstypen

```

typedef enum {
 Mon, Tue, Wed, Thu, Fri, Sat, Sun
} day;

data Day = Mon | Tue | Wed | Thu | Fri | Sat | Sun
Ü: enum in Java

```

### Designfragen:

- automatische oder manuelle Konversion zw. Aufzählungstypen (und zugrundeliegendem Zahltyp)

### Maßeinheiten in F#

- physikalische Größe = Maßzahl × Einheit.
- viele teure Softwarefehler durch Ignorieren der Einheiten.
- in F# (Syme, 200?), aufbauend auf ML (Milner, 197?)

```

[<Measure>] type kg ;; let x = 1<kg> ;;
x * x ;;
[<Measure>] type s ;; let y = 2<s> ;;
x * y ;; x + y ;;

```

- <https://docs.microsoft.com/en-us/dotnet/fsharp/language-reference/units-of-measure>

### Zeichen und Zeichenketten

- das naive Modell ist:
  - Zeichen paßt in (kurze) Maschinenzahl (z.B. char = byte)
  - Zeichenketten sind (Zeiger auf) Arrays

das ist historisch begründet (US-amerikanische Hardware-Hersteller, lateinisches Alphabet)

- das umfassende Modell ist <http://www.unicode.org/versions/Unicode12.1.0/> (insbes. Kapitel 2)  
jedes Zeichen wird durch *encoding scheme* (z.B. UTF8) auf *Folge* von code units (z.B. Bytes) abgebildet.

## Zusammengesetzte Typen

Typ = Menge, Zusammensetzung = Mengenoperation:

- Produkt (record, struct)
- disjunkte Summe (union, case class, enum)
- Rekursion, z.B.

```
data List a = Nil | Cons a (List a)
```

- Potenz (Funktion), z.B.

```
type Sorter a = (List a -> List a)
```

## Produkttypen (Records)

- $R = A \times B \times C$
- Kreuzprodukt mit benannten Komponenten:

```
typedef struct {
 A foo; B bar; C baz;
} R;
R x; ... B y = x.bar; ...
```

- erstmalig in COBOL ( $\leq 1960$ ) (Interview mit der Erfinderin: [http://archive.computerhistory.org/resources/text/Oral\\_History/Hopper\\_Grace/102702026.05.01.pdf](http://archive.computerhistory.org/resources/text/Oral_History/Hopper_Grace/102702026.05.01.pdf))

## Summen-Typen

- $R = A \cup B \cup C$
- disjunkte (diskriminierte) Vereinigung (Pascal)

```
type tag = (eins, zwei, drei);
type R = record case t : tag of
 eins : (a_value : A);
 zwei : (b_value : B);
 drei : (c_value : C);
end record;
```

- nicht diskriminiert (C):

```
typedef union {
 A a_value; B b_value; C c_value;
} R;
```

### Vereinigung mittels Interfaces

*I* repräsentiert die Vereinigung von *A* und *B*:

```
interface I { }
class A implements I { int foo; }
class B implements I { String bar; }
```

Notation dafür in Scala (<http://scala-lang.org/>)

```
abstract class I
case class A (foo : Int) extends I
case class B (bar : String) extends I
```

### Verarbeitung durch *Pattern matching*

```
def g (x : I): Int = x match {
 case A(f) => f + 1
 case B(b) => b.length() }
```

### Rekursive algebraische Datentypen

Haskell (<http://haskell.org/>)

```
data Tree a = Leaf a
 | Branch (Tree a) (Tree a)
```

### Java

```
interface Tree<A> { }
class Leaf<A> implements Tree<A> { A key }
class Branch<A> implements Tree<A>
{ Tree<A> left, Tree<A> right }
```

*Tree a* ist ein *algebraischer Datentyp*:

- die Konstruktoren (Leaf, Branch) bilden die *Signatur* der Algebra,
- die *Elemente* der Algebra sind Terme (Bäume)

## Potenz-Typen

- $B^A := \{f : A \rightarrow B\}$  (Menge aller Funkt. von  $A$  nach  $B$ )
- Potenz ist sinnvolle Notation, denn  $|B|^{|A|} = |B^A|$
- Realisierungen:
  - Funktionen (Unterprogramme)
  - Wertetabellen (Funktion mit endlichem Definitionsbereich) (Assoziative Felder, Hashmaps)
  - Felder (Definitionsbereich ist Aufzählungstyp) (Arrays)
  - Zeiger (Hauptspeicher als Array)
  - Zeichenketten (Strings)
- die unterschiedliche Notation dafür ist bedauerlich.

```
f(42); f.get(42); f[42]; *(f+42); f.charAt(42)
```

## Felder (Arrays)

- Realisierung einer Abbildung, Definitionsbereich ist Intervall von Zahlen, Wertebereich ist benutzerdefiniert.
- Motivation: Zugriff auf beliebiges Element in konstanter Zeit (unabhängig von Intervallgröße)

```
a[i] = * (a + w * i)
```

- Design-Entscheidungen:
  - welche Index-Typen erlaubt? (Zahlen? Aufzählungen?)
  - Bereichsprüfungen bei Indizierungen? (C:nein, Java:ja)
  - Allokation statisch oder dynamisch?
  - Index-Bereiche statisch oder dynamisch?
  - mehrdimensionale Felder (gemischt oder rechteckig)?

## Felder in C

```
int main () {
 int a [10][10];
 a[3][2] = 8;
 a[2][12] = 5;
 printf ("%d\n", a[3][2]);
}
```

- statische Dimensionierung,
- dynamische Allokation,
- keine Bereichsprüfungen.
- Form: rechteckig, Adress-Rechnung:

```
int [M][N]; a[x][y] ==> *(&a + (N*x + y))
```

## Felder in Javascript

- die Notation `a[i]` wird verwendet für Felder (Zugriff über Index) *und* (Hash)Maps (Zugriff über Schlüssel).
- durch das Fehlen statischer Typisierung sowie implizite Umwandlung zwischen Zahl und Zeichenkette wird absurdes Verhalten spezifiziert, vgl. <https://news.ycombinator.com/item?id=14675706>

```
var arr1 = []; arr1[4294967296]=1;
// arr1.length == 0
var arr2 = []; arr2[2147483647]=1;
// arr2.length == 2147483648
var arr3 = []; arr3[-1]=1;
// arr3.length == 0
```

## Felder in Java

```
int [][] feld =
 { {1,2,3}, {3,4}, {5}, {} };
for (int [] line : feld) {
 for (int item : line) {
 System.out.print (item + " ");
 }
 System.out.println ();
}
```

- dynamische Dimensionierung und Allokation,
- Bereichsprüfungen.
- Arrays sind immer eindimensional, aber man kann diese schachteln. (Kosten?)

### Kosten der Bereichsprüfungen

- es wird oft als Argument für C (und gegen Java) angeführt, daß die erzwungene Bereichsprüfung bei jedem Array-Zugriff so teuer sei.
- sowas sollte man erst glauben, wenn man es selbst gemessen hat.
- moderne Java-Compiler sind *sehr clever* und können *theorem-prove away (most) subscript range checks*
- das kann man auch in der Assembler-Ausgabe des JIT-Compilers sehen.  
<http://www.imn.htwk-leipzig.de/~waldmann/etc/safe-speed/>

### Felder in C#

Übung: Unterschiede zwischen

- `int [][] a` geschachtelt (wie in Java)
- `int [,] a` mehrdimensional rechteckig

in

- Benutzung (Zugriff)
- Konstruktion/Initialisierung

### Verweistypen

- Typ  $T$ , Typ der Verweise auf  $T$ .
- Operationen: new, put, get, delete
- ähnlich zu Arrays (das Array ist der Hauptspeicher)
- explizite Verweise in C, Pascal

```
int x = 2 ; int *p = &x; ... *p + 3
```

- implizite Verweise: Java:  
alle nicht primitiven Typen sind Verweistypen, De-Referenzierung ist implizit  
Object a = ...; Object b = a; kopiert Verweis
- C#: class ist Verweistyp, struct ist Werttyp

### Verweis- und Wertsemantik in C#

- für Objekte, deren Typ `class ...` ist:  
Verweis-Semantik (wie in Java)
- für Objekte, deren Typ `struct ...` ist:  
Wert-Semantik

Testfall:

```
class s {public int foo; public string bar;}
s x = new s(); x.foo = 3; x.bar = "bar";
s y = x; y.bar = "foo";
Console.WriteLine (x.bar);
```

und dann `class` durch `struct` ersetzen

### Algebraische Datentypen in Pascal, C

Rekursion unter Verwendung von Verweistypen

Pascal:

```
type Tree = ^ Node ;
type Tag = (Leaf, Branch);
type Node = record case t : Tag of
 Leaf : (key : T) ;
 Branch : (left : Tree ; right : Tree);
end record;
```

C: ähnlich, benutze `typedef`

## Null-Zeiger: der Milliarden-Dollar-Fehler

- Tony Hoare (2009): [The null reference] has led to innumerable errors, vulnerabilities, and system crashes, which have probably caused a billion dollars of pain and damage in the last forty years.

(<https://www.infoq.com/presentations/Null-References-The-Billion-Doll>)

- Das Problem sind nicht die Zeiger selbst, sondern daß (in vielen Sprachen) der Wert `null` zu jedem Zeigertyp gehört — obwohl er gar kein Zeiger ist.

Das ist die Verwechslung zwischen `t` und `Maybe t`.

## Hausaufgaben Typen

1. für Mengen  $A = \emptyset, B = \{0\}, C = \{1, 2\}, D = \{3, 4, 5\}, E = \{6, 7, 8, 9\}$ , geben Sie an:

- alle Elemente von  $A \times C, B \times D, A \cup B, B^A, A^B, C^B, B^C, C^D$
- ein Element aus  $(C \times D)^E$
- die Kardinalitäten von  $(C \times D)^E, C^{D \cup E}$

2. zur Folie „Felder in C“:

Programm kompilieren, ausführen.

Assembler-Code ausgeben und erklären (`gcc -S` oder `clang -S`)

3. Welche offiziellen Pläne gibt es für Wert-Typen in Java?

Anhand von Original-Quellen (!) belegen, ggf. Beispiele vorführen, falls experimentelle Implementierung vorhanden.

4. Welche offiziellen Pläne gibt es für Pattern Matching (ähnlich zu algebraischen Datentypen, nicht verwechseln mit Matching für reguläre Ausdrücke) in ECMA-Script (JS)

Anhand von Original-Quellen (!) belegen, ggf. Beispiele vorführen. falls experimentelle Implementierung vorhanden.

## 6 Bezeichner, Bindungen, Bereiche

### Variablen

- vereinfacht: Variable bezeichnet eine Speicherstelle
- genauer: Variable besitzt Attribute
  - Name
  - Adresse
  - Wert
  - Typ
  - Lebensdauer
  - Sichtbarkeitsbereich
- Festlegung dieser Attribute *statisch* oder *dynamisch*

### Namen in der Mathematik

- ein Name bezeichnet einen unveränderlichen Wert
$$e = \sum_{n \geq 0} \frac{1}{n!}, \quad \sin = (x \mapsto \sum_{n \geq 0} (-1)^n \frac{x^{2n+1}}{(2n+1)!})$$
- auch  $n$  und  $x$  sind dabei lokale Konstanten (werden aber gern „Variablen“ genannt)
- auch die „Variablen“ in Gleichungssystemen sind (unbekannte) Konstanten  $\{x + y = 1 \wedge 2x + y = 1\}$

in der Programmierung:

- Variable ist Name für Speicherstelle (= konstanter Zeiger)
- implizite Dereferenzierung beim Lesen und Schreiben
- Konstante: Zeiger auf schreibgeschützte Speicherstelle

### Namen

- welche Buchstaben/Zeichen sind erlaubt?
- reservierte Bezeichner?

- Groß/Kleinschreibung?
- Konvention: `long_name` oder `longName` (camel-case)  
(Fortran: `long name`)  
im Zweifelsfall: Konvention der Umgebung einhalten
- Konvention: Typ im Namen (Bsp.: `myStack = ...`)
  - verrät Details der Implementierung
  - ist ungeprüfte Behauptung

besser: `Stack<Ding> rest_of_input = ...`

## Deklaration und Definition

- Bsp: `int x = 8;` `int x` ist Deklaration, `= 8` ist Definition
- Bsp: `static int f(int y) { return y+1; }` `static int f(int y)` ist Deklaration, `(int y) { return y+1; }` ist Definition.
- Deklaration:
  - statische Semantik: der Name ist ab hier sichtbar
  - dynamische S.: dem Namen ist Speicherplatz zugeordnet
- Definition:
  - dynamische Semantik: dem Namen ist Wert zugeordnet
  - statische S.: (siehe „garantierte Initialisierung“ später)

## Typen für Variablen

- dynamisch (Wert hat Typ)
- statisch (Name hat Typ)
  - deklariert (durch Programmierer)
  - inferiert (durch Übersetzer)  
z. B. `var` in C#

- Vor/Nachteile: Lesbarkeit, Sicherheit, Kosten  
der Typ eines Bezeichners ist seine beste Dokumentation  
(weil sie maschinell überprüft wird - bei statischer Typisierung)

### **Dynamisch typisierte Sprachen**

- Daten sind typisiert, Namen sind nicht typisiert.
- LISP, Clojure, PHP, Python, Perl, Javascript, ...
- ```
var foo = function(x) {return 3*x;};  
foo(1);  
foo = "bar";  
foo(1);
```

Statisch typisierte Sprachen

- Namen sind typisiert, Daten sind typisiert (? siehe unten)
- Invariante:
zur Laufzeit ist der *dynamische Typ* des Namens (der Typ des Datums auf der durch den Namen bezeichneten Speicherstelle)
immer gleich dem *statischen Typ* des Namens
- woher kommt der statische Typ?
 - Programmierer deklariert Typen von Namen (C, Java)
 - Compiler inferiert Typen von Namen (ML, C# (var))
- dynamischer Typ muß zur Laufzeit nicht repräsentiert werden (das spart Platz u. Zeit): Compiler erzeugt Code, der das Resultat der Laufzeitprüfung vorwegnimmt.

Programmablaufsteuerung

- auch in statisch typisierten Sprachen:
Programm-Ablauf soll von (Eingabe)daten abhängen
(nicht bei jedem Programm, vgl. Matrix-Multiplikation)

- Bsp. (Haskell) Fallunterscheidung nach Konstruktor

```
data T = Foo ... | Bar ...
case a of { Foo x -> ... ; Bar y -> ... }
```

in OO: Dispatch nach implementierender Klasse

- dabei wird Laufzeit(typ)Information (RTTI) benutzt
diese kann effizient repräsentiert werden
(für Auswahl aus 2 Konstruktoren: reicht 1 Bit)

Typdeklarationen

- im einfachsten Fall (Java, C#):

```
Typname Variablenname [ = Initialisierung ] ;
int [] a = { 1, 2, 3 };
Func<double,double> f = (x => sin(x));
```

- gern auch komplizierter (C): dort gibt es keine Syntax für Typen, sondern nur für Deklarationen von Namen.

```
double f (double x) { return sin(x); }
int * p; double ( * a [2]) (double) ;
```

Beachte: * und [] werden „von außen nach innen“ angewendet

- Ü: Syntaxbäume zeichnen, a benutzen

Typinferenz in C# und Java

- C#

```
public class infer {
    public static void Main (string [] argv) {
        var arg = argv[0];
        var len = arg.Length;
        System.Console.WriteLine (len);
    }
}
```

- Ü: das `var` in C# ist nicht das `var` aus Javascript.
- Java:
für formale Parameter von anonymen Unterprogrammen

```
Function<Integer,Integer> f = (x) -> x;
```

Code-Inferenz

- in vielen einfachen Sprachen dienen Typen tatsächlich „nur“ zur Spezifikation und Dokumentation
... man könnte sie also doch weglassen, wenn man nur die Implementierung selbst braucht?
- moderne, ausdrucksstarke Typsysteme nützen deutlich mehr und tragen auch zur *Code-Erzeugung* bei.
Sandy Maguire: <https://thinkingwithtypes.com/>
- Anwendungen/Beispiele (u.a. in autotool)
 - typgesteuerte Testdatengenerierung <https://hackage.haskell.org/package/leancheck>
 - *Type-Level Web APIs with Servant*, Alp Mestanogullari et al., 2015, <https://www.servant.dev/>

Konstanten

- = Variablen, an die genau einmal zugewiesen wird
- – C: `const` (ist Attribut für Typ)
– Java: `final` (ist Attribut für Variable)
- Vorsicht:

```
class C { int foo; }  
static void g (final C x) { x.foo ++; }
```
- alle Deklarationen so lokal und so konstant wie möglich!
(d. h., Attribute *immutable* usw.)
denn das verringert den Umfang der Dinge, über die man nachdenken muß, um das Programm zu verstehen

Lebensort und -Dauer von Name und Daten

- statisch (auf statisch zugeordneter Adresse im Hauptspeicherbereich)

```
int f (int x) {  
    static int y = 3; y++; return x+y; }
```

- dynamisch (auf zur Laufzeit bestimmter Adresse)
 - Stack (Speicherbereich für Unterprogramm-Aufruf) { int x = ... }
 - Heap (Hauptspeicherbereich)
 - * explizit (new/delete, malloc/free)
 - * implizit (kein delete, sondern automatische Freigabe)
- Beachte (in Java, C#) in { C x = new C (); } ist x stack-lokal, Inhalt ist Zeiger auf das heap-globale Objekt.

Sichtbarkeit von Namen

- eine Deklaration ist sichtbar, wenn die Verwendung des Namens ein Bezug auf die deklarierte Variable ist
- üblich ist: Sichtbarkeit beginnt nach Deklaration und endet am Ende des umgebenden Blockes.
- Import-Deklarationen machen Namen aus anderen Namensbereichen sichtbar
- (Java) ohne Import-D. besteht *qualifizierte* Sichtbarkeit
- (C): Sichtbarkeit beginnt in der Initialisierung

```
int x = sizeof(x); printf ("%d\n", x);
```

Ü: ähnliches Beispiel für Java? Vgl. JLS Kapitel 6.

Verdeckung von Deklarationen

- Namen sind auch in inneren Blöcken sichtbar:

```
int x;
while (..) {
  int y; ... x + y ...
}
```

- innere Deklarationen verdecken äußere:

```
int x;
while (..) {
  int x; ... x ...
}
```

Sichtbarkeit in JavaScript

- Namen sind sichtbar
 - Deklaration mit `var`: im (gesamten!) Unterprogramm


```
(function() { { var x = 8; } return x; } ) ()
```
 - Deklaration mit `let`: im (gesamten!) Block


```
(function() { { let x = 8; } return x; } ) ()
```
- Ü: erkläre (durch Verweis auf Sprachspezifikation)

```
(function(){let x=8; {x=9} return x} )()
(function(){let x=8; {x=9;let x=10} return x} )()
(function(){let x=8; {y=9;let x=10} return x} )()
```

Übung

- Wo steht im Java-Sprachstandard, daß jeder Name vor seiner Benutzung deklariert sein muß?
- Beschreiben Sie Sichtbarkeitsbereich und Lebensdauer der Variablen in

```
int f (int x) {
  int a = x;
  int b = a+1;
  int c = b+2;
  return c;
}
```

Vergleichen Sie mit der Ausgabe (Assemblercode) von `gcc -S`

- Beispiel in Java für: Sichtbarkeit beginnt mit Initialisierung

Hausaufgaben

1. Beobachten und erklären Sie die Ausgabe von

```
#include <stdio.h>
int main (int argc, char **argv) {
    int x = 3;
    { printf ("%d\n", x);
      int x = 4;
      printf ("%d\n", x);
    }
    printf ("%d\n", x);
}
```

schreiben Sie ein entsprechendes Java-Programm und vergleichen Sie.

2. Sichtbarkeit von Deklarationen in Javascript.

Siehe Folie, Original-Dokumentation zeigen und Beispiele vorführen (node), ergänzen durch weitere Beispiele mit nicht offensichtlicher Semantik, Bsp: Variablen-Deklaration in einem Zweig einer Verzweigung.

nur Sichtbarkeiten — Programmablaufsteuerung soll trivial sein (keine Schleifen, keine Unterprogramme)

3. die Aufgabe passend zur Jahreszeit:

Bearbeiten Sie eine der aktuellen Aufgaben aus dem DMV-Adventskalender <https://www.mathekalender.de/index.php?page=calendar>

mit Hilfe eines Programm in einer esoterischen Sprache (d.h., nicht C, C++, C, Java, JS, Haskell, Python u.ä. aus Pflicht-Lehrveranstaltungen)

Beispiele für esoterische Sprachen: <https://web.archive.org/web/20200804145505/http://www.99-bottles-of-beer.net/>

Sprache soll eine Open-Source-Implementierung haben (Compiler, Interpreter), die Sie selbst installieren, im Linux-Pool oder vergleichbar. (Also nicht: Webinterface, binäre Download.)

Programm soll eine deutliche Beziehung zur Kalender-Aufgabe haben. (Also nicht: einfach ein Demo der Sprache kopieren.) Programm muß die jeweilige Aufgabe nicht vollautomatisch lösen. Es könnte z.B. Experimente automatisieren, damit man besser sieht, was eigentlich zu beweisen ist.

Methoden der Beschreibung von Programmiersprachen (aus unserer Vorlesung) sollen angewendet werden.

Ab jetzt jede Woche möglich, jedesmal max. 2 Gruppen.

Insgesamt (bis Semester-Ende) soll keine Sprache doppelt vorkommen, deswegen bitte vorher untereinander absprechen.

7 Ausdrücke

Definition, Abgrenzung

- Ausdruck hat *Wert* (Zahl, Objekt, ...)
(Ausdruck wird *ausgewertet*)
- Anweisung hat *Wirkung* (Änderung des Speicher/Welt-Zustandes)
(Anweisung wird *ausgeführt*)

Vgl. Trennung (in Pascal, Ada)

- Funktion (Aufruf ist Ausdruck)
- Prozedur (Aufruf ist Anweisung)

Ü: wie in Java ausgedrückt? wie stark getrennt?

Syntax von Ausdrücken

- einfache Ausdrücke : Literale, (Variablen-)Namen
- zusammengesetzte Ausdrücke:
 - Operator-Symbol zwischen Argumenten
 - Funktions-Symbol vor Argument-Tupel

wichtige Spezialfälle für Operatoren:

- arithmetische (von Zahlen nach Zahl)

- relationale (von Zahlen nach Wahrheitswert)
- boolesche (von Wahrheitswerten nach Wahrheitsw.)

Wdhlg: Syntaxbaum, Präzedenz, Assoziativität.

Designfragen für Ausdrücke

- Syntax
 - Präzedenzen (Vorrang)
 - Assoziativitäten (Gruppierung)
 - kann Programmierer neue Operatoren definieren?
- statische Semantik
 - ... vorhandene Operatornamen überladen?
 - Typen der Operatoren?
 - implizite, explizite Typumwandlungen?
- dynamische Semantik
 - Ausdrücke dürfen (Neben-)Wirkungen haben?
 - falls mehrere: in welcher Reihenfolge?

Beziehungen zw. Ausdruck und Anweisung

- in allen imperativen Sprachen gibt es Ausdrücke mit Nebenwirkungen (nämlich Unterprogramm-Aufrufe)
- in den rein funktionalen Sprachen gibt es keine (Neben-)Wirkungen, also keine Anweisungen (sondern nur Ausdrücke).
- in den C-ähnlichen Sprachen ist = ein Operator, (d. h. die Zuweisung ist syntaktisch ein Ausdruck, kann Teil von anderen Ausdrücken sein)


```
int x = 3; int y = x + (x = 4);
```
- in den C-ähnlichen Sprachen:

Ausdruck ist als Anweisung gestattet (z.B. in Block)

```
{ int x = 3; x++ ; System.out.println(x); }
```

Überladene Operatornamen

- Def: Name n *überladen*, falls n mehrere Bedeutungen hat
- in vielen Sprachen sind arithmetische und relationale Operatornamen überladen ...
weil das Typsystem keine flexiblere Lösung gestattet, wie z.B. `class Num a where (+) :: a -`
- Überladung wird statisch aufgelöst (vom Compiler, anhand der Typen der Argument-Ausdrücke)
- `int x = 3; int y = 4; ... x + y ...`
`double a; double b; ... a + b ...`
`String p; String q; ... p + q ...`

Der Plus-Operator in Java

- ... addiert Zahlen und verkettet Strings.
- `System.out.println ("foo" + 3 + 4);`
`System.out.println (3 + 4 + "bar");`
- Vorgehen für die Analyse:
 - abstrakten Syntaxbaum bestimmen
 - Typen (als Attribute der AST-Knoten) bestimmen,
 - dabei implizite Typ-Umwandlungen einfügen
(in diesem Fall `Integer.toString()`)
 - Werte (als Attribute) bestimmen

Automatische Typanpassungen

- in vielen Sprachen postuliert man eine Hierarchie von Zahlbereichstypen:
`byte \subseteq int \subseteq float \subseteq double`
im allgemeinen ist das eine Halbordnung.
- Operator mit Argumenten verschiedener Typen: `(x :: int) + (y :: float)`
beide Argumente werden zu kleinstem gemeinsamen Obertyp promoviert, falls dieser eindeutig ist (sonst statischer Typfehler)
(Halbordnung \rightarrow Halbverband)
- (das ist die richtige Benutzung von *promovieren*)

Implizite/Explizite Typumwandlungen

- Was druckt dieses Programm?

```
long x = 1000 * 1000 * 1000 * 1000;
long y = 1000 * 1000;
System.out.println ( x / y );
```

- Was druckt dieses Programm?

```
System.out.println ((int) (char) (byte) -1);
```

- Moral: wenn man nicht auf den ersten Blick sieht, was ein Programm macht, dann macht es wahrscheinlich nicht das, was man will.

Explizite Typumwandlungen

sieht gleich aus und heißt gleich (cast), hat aber verschiedene Bedeutungen:

- Datum soll in anderen Typ gewandelt werden, Repräsentation ändert sich:

```
int x = 4; double y = (double) x / 5;
/* Ü : */ double z = (double) (x / 5) ;
```

- Programmierer weiß es besser als der Compiler, Code für Typprüfung zur Laufzeit wird erzeugt, Repräsentation ändert sich nicht:

```
List books; Book b = (Book) books.get (7);
```

Typumwandlungen in Haskell

- Joachim Breitner et al.: *Safe zero-cost coercions for Haskell*, JFP 2016, <https://dblp.org/rec/journals/jfp/BreitnerEJW16.html>

Umwandlung zwischen Basistyp und abgeleitetem Typ mit gleicher Laufzeit-Repräsentation

- sicher: durch Compiler bewiesen
- kostenlos: keine Laufzeitkosten

```
newtype Foo = Foo Int
data Bar a = Bar Bool a
xs = replicate 10 (Bar True (Foo 3)) :: [Bar Foo]
ys = Data.Coerce.coerce xs :: [Bar Int]
```

Der Zuweisungs-Operator

- Syntax:
 - Algol, Pascal: Zuweisung `:=`, Vergleich `=`
 - Fortran, C, Java: Zuweisung `=`, Vergleich `==`
- Semantik der Zuweisung `a = b`:
 - bestimme Adresse (lvalue) p von a
 - bestimme Wert (rvalue) v von b
 - schreibe v auf p
- diese Ausdrücke haben einen lvalue:
 - Variablen
 - $a[i]$, mit: rvalue von a ist Array, rvalue von i ist Index
 - $o.a$, mit: rvalue von o ist Objekt mit Attribut a

Bsp: `foo() [bar()]`

Weitere Formen der Zuweisung

(in C-ähnlichen Sprachen)

- verkürzte Zuweisung: `a += b`
entsprechend für andere binäre Operatoren
 - lvalue p von a wird bestimmt (nur einmal)
 - rvalue v von b wird bestimmt
 - Wert auf Adresse p wird um v erhöht
- Inkrement/Dekrement
 - Präfix-Version `++i`, `--j`: Wert ist der geänderte
 - Suffix-Version `i++`, `j--`: Wert ist der vorherige
- Ü: experimentell bestätigen, daß lvalue des Zuweisungsziels nur einmal ausgewertet wird

Teil-Ausdrücke mit Nebenwirkungen

(*side effect*; falsche Übersetzung: Seiteneffekt)

- in C-ähnlichen Sprachen: Zuweisungs-Operatoren bilden Ausdrücke, d. h. Zuweisungen sind Ausdrücke und können als Teile von Ausdrücken vorkommen.
- Wert einer Zuweisung ist der zugewiesene Wert

```
int a; int b; a = b = 5; // wie geklammert?
```

- Komma-Operator zur Verkettung von Ausdrücken (mit Nebenwirkungen) – vgl. C mit Java

```
for (... ; ... ; i++, j--) { ... }
```

Auswertungsreihenfolgen

- zusammengesetzte Programme können mehrere Bestandteile mit Nebenwirkungen haben.

In welcher Reihenfolge finden diese statt?

- Anweisungen: explizite Programm-Ablauf-Steuerung

```
{ int a = 5; a = 6; int b = a + a; }
```

- Ausdrücke?

```
{ int a; int b = (a = 5) + (a = 6); }
```

- C, C++: Reihenfolge nicht spezifiziert, wenn Wert davon abhängt, dann ist Verhalten *nicht definiert*
- Java, C#: Reihenfolge genau spezifiziert (siehe JLS)

Ausdrucks-Semantik von C

- Sprachstandard benutzt Begriff *sequence point* (Meilenstein): bei Komma, Fragezeichen, && und ||
- Nebenwirkungen zwischen Meilensteinen müssen *unabhängig* sein (nicht die gleiche Speicherstelle betreffen),

- ansonsten ist das Verhalten *undefiniert*, d.h., der Compiler darf *beliebigen* Code erzeugen, z.B. solchen, der die Festplatte löscht oder Cthulhu heraufbeschwört.
- vgl. Aussagen zu sequence points in <http://gcc.gnu.org/readings.html> und Gurevich, Huggins: *Semantics of C*, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.29.6755>

Der ternäre Verzweigungs-Operator ?:

- ```
if (0 == x % 2) { x = x / 2; }
else { x = 3 * x + 1; }
```

```
x = if (0 == x % 2) { x / 2 }
 else { 3 * x + 1 } ;
```

- historische Notation dafür benutzt ternären Operator ?:

```
x = (0 == x % 2) ? x / 2 : 3 * x + 1;
```

- Verzweigungs-Operator auf lvalues (C++):

```
int a = 4; int b = 5; int c = 6;
(c < 7 ? a : b) = 8;
```

### Logische (Boolesche) Ausdrücke

- Konjunktion &&, Disjunktion ||, Negation !  
Äquivalenz, Antivalenz
- *verkürzte Auswertung* für Konjunktion und Disjunktion: wenn nach Auswertung des linken Arguments das Resultat feststeht, denn werte rechtes nicht aus

```
int [] a = ...; int k = ...;
if (k >= 0 && a[k] > 7) { ... }
```

- warum keine verkürzte Auswertung für Äquiv., Antiv.?
- Ü: für die Sprache C:

```
if (3 && 4) ... ; if (3 & 4) ...
```

## Relationale Operatoren

- kleiner, größer, gleich,...
- Übung: Was tut dieses Programm (in C? in Java?)

```
int a = -4; int b = -3; int c = -2;
if (a < b < c) {
 printf ("aufsteigend");
}
```

## Übungen

1. Gary Bernhardt: WAT (2012) <https://www.destroyallsoftware.com/talks/wat>
2. über die implizite Konversion von `int` nach `float`,
  - (a) Es gilt nicht  $\text{int} \subseteq \text{float}$ , denn beide Mengen sind gleich groß (wie groß?) und es gibt (viele)  $y \in \text{float} \setminus \text{int}$  (welche?)
  - (b) Geben Sie ein  $x \in \text{int} \setminus \text{float}$  explizit an.
  - (c) Wo ist diese Konversion in der Sprachspezifikation beschrieben?
  - (d) desgleichen für `long` zu `double`
  - (e) Gilt  $\text{int} \subseteq \text{double}$ ?
    - a) laut Spezifikation, b) wirklich?
3. Wiederholung Operator-Syntax:
  - ist die Mengendifferenz assoziativ?
  - vgl. <https://gitlab.haskell.org/ghc/ghc/issues/15892> „the fix was to add a pair of parentheses“
4. Wiederholung: Was spricht dafür und dagegen, daß in einem Programmtext neue Operatoren definiert werden?

In C++ darf man keine neuen Operatoren deklarieren, aber vorhandene Operatoren neu implementieren. Begründen Sie diese Design-Entscheidung.
5. durch Verweis auf JLS erklären:

- `System.out.println ( 12345 + 54321 );`
- `System.out.println ("H" + "a");`  
`System.out.println ('H' + 'a');`
- `char x = 'X'; int i = 0;`  
`System.out.print (true ? x : 0);`  
`System.out.print (false ? i : x);`

Quelle: Joshua Bloch, Neil Gafter: *Java Puzzlers*, Addison-Wesley, 2005.

## Hausaufgaben

1. UB (undefined behaviour) für C-Ausdrücke mit abhängigen Teilausdrücken zwischen Sequence Points:
  - (a) Ausgehend von Beispielen auf Folie: finden Sie Programme, für die
    - verschiedene Compiler (gcc, clang)
    - ein Compiler (gcc) bei verschiedenen Optionen (`-O0`, `-O3`, ...)
    - verschiedene Versionen eines Compilers (im Pool: verschiedene gcc)
 Code mit unterschiedlichem Verhalten erzeugen.
  - (b) Vergleichen Sie mit Semantik des entsprechenden Java-Programms. (Ausführen, Bytecode, Sprachspezifikation)
  - (c) Wer ist Cthulhu? Seine Beziehung zu Semantik von C-Programmen ist Folklore, Belege z.B. „after all, we have UB here, and this could just summon cthulhu or anything else“.
2. Verkürzte Auswertung bei logischen Operatoren in Java und JS (Tests mit `jshell`, `node`)
  - (a) einen Testfall angeben, der die verkürzte Auswertung bei `||` zeigt.
  - (b) Der Operator `|` verknüpft Zahlen bitweise. (Testfall angeben)  
Es gibt `|` auch für `boolean`. Worin besteht der Unterschied zu `||` ? (Testfall angeben)
  - (c) desgl. für `&`
  - (d) das gleiche für JS oder TS untersuchen

3. Verkürzte Auswertung bei logischen Operatoren in Ada: Sprachstandard und Vorführung. Benutze GNAT (GNU Ada Translator) als Teil von GCC (GNU Compiler Collection), ist im Pool installiert
4. (optional) DMV-Advents-Kalender-Aufgabe(n)

## 8 Anweisungen

### Definition

- Semantik: Ausführen einer Anweisung bewirkt Zustandsänderung
- abstrakte Syntax:
  - einfache Anweisung:
    - \* leere Anweisung (*skip*), Zuweisung ( $l := r$ ),
    - \* Sprung *goto*, *break*, *continue*, *return*, *throw*
    - \* Unterprogramm-Aufruf ( $p(a, b)$ )
  - zusammengesetzte Anweisung:
    - \* Nacheinanderausführung (Block)
    - \* Verzweigung (zweifach: *if*, mehrfach: *switch*)
    - \* Wiederholung (Schleife)

### Programm-Ablauf-Steuerung

- engl. *control flow*, falsche Übersetzung: Kontrollfluß;  
*to control* = steuern, *to check* = kontrollieren/prüfen
- von-Neumann-Modell: jede Anweisung beschreibt
  - Was? (Operation) Womit? (Operanden) Wohin? (Resultat)
  - Wie weiter? (nächste Anweisung)

Programm-Ablauf dabei also durch Sprünge gesteuert

- Ablaufsteuerung durch strukturierte Programmierung:  
jedes Teilprogramm (Teilbaum des AST)  
hat genau einen Eingang und genau einen Ausgang
- vorzeitiges Verlassen eines Teilprogramms:  
Schleife (*break*, *continue*), UP (*return*), *throw/catch*

## Blöcke

- Def: Folge von (Deklarationen und) Anweisungen
- Designfrage/historische Entwicklung: Deklarationen ...
  - am Beginn des Progr. (Assembler, COBOL, Fortran)
  - am Beginn jedes Unter-Programms (Pascal)
  - am Beginn jedes Blocks (C)
  - an jeder Stelle jedes Blocks (C++, Java)
- Designfrage für Syntax: Blöcke
  - explizit (Klammern, begin/end)
  - implizit (`if ... then ... end if`, d.h., ohne begin)

## Verzweigungen (zweifach)

- in vielen Sprachen:

```
if Bedingung then Anweisung1
 [else Anweisung2]
```
- Designfrage (Syntax und Semantik): Bedingung ist ...
  - beliebiger Ausdruck mit Typ Wahrheitswert
  - nur Vergleich zwischen Ausdrücken vom Typ Zahl
- Designfrage Syntax: Mehrdeutigkeit der Grammatik
  - gelöst durch Festlegung (else gehört zu letztem if)
  - vermieden durch Block-Bildung (Ada)
  - tritt nicht auf, weil man `else` nie weglassen darf, weil beide Zweige einen Wert liefern (`? :`, Haskell)

## Mehrfach-Verzweigung

Syntax:

```
switch (e) {
 case c1 : s1 ;
 case c2 : s2 ;
 [default : sn;] }
```

Semantik

```
if (e == c1) s1
else if (e == c2) s2
... else sn
```

- Bezeichnung: der Ausdruck  $e$  heißt *Diskriminante*
- Vorsicht! Das ist *nicht* die Semantik in C(++), Java.
- welche Typen für  $e$ ? (z.B.: Aufzählungstypen)
- Wertebereiche? (case  $c_1 \dots c_2 : \dots$ )
- was passiert, wenn mehrere Fälle zutreffen?  
(z.B.: statisch verhindert dadurch, daß  $c_i$  verschiedene Literale sein müssen)

## switch/break

```
switch (index) {
 case 1 : odd ++;
 case 2 : even ++;
 default :
 printf ("wrong index %d\n", index);
}
```

- Semantik in C, C++, Java ist nicht „führe den zum Wert der Diskriminante passenden Zweig aus“
- sondern „...passenden Zweig aus *sowie alle danach folgenden Zweige*“.
- C#: jeder Zweig *muß* mit `break` oder `goto` enden.

## Kompilation

ein switch (mit vielen cases) wird übersetzt in:

- (naiv) eine lineare Folge von binären Verzweigungen (if, elsif)
- (semi-clever) einen balancierter Baum von binären Verzweigungen
- (clever) eine Sprungtabelle

## Übung:

- einen langen Switch (1000 Fälle) erzeugen (durch ein Programm!)
- Assembler/Bytecode anschauen

## Pattern Matching

- Fallunterscheidung nach dem Konstruktor
- Bindung von lokalen Namen

```
abstract class Term // Scala
case class Constant (value : Int)
 extends Term
case class Plus (left: Term, right : Term)
 extends Term
def eval(t: Term): Int = {
 t match {
 case Constant(v) => v
 case Plus(l, r) => eval(l) + eval(r)
 } }
}
```

## Wiederholungen

- Programmablaufsteuerung von Wiederholungen:
  - von-Neumann-Modell (Maschine, Assembler): unbedingter, bedingter Sprung
  - strukturierte Programmierung: Schleifen
- Designfragen für Schleifen:
  - wie wird Schleife gesteuert?  
Bedingung, Zähler, Zustand (Iterator), Daten (Collection)
  - an welcher Stelle in der Schleife findet Steuerung statt (Anfang, Ende, dazwischen, evtl. mehrere Stellen)

## Schleifen steuern durch...

- Zähler

```
for p in 1 .. 10 loop .. end loop;
```

- Daten

```
map (\x -> x*x) [1,2,3] ==> [1,4,9]
```

```
Collection<String> c
 = new LinkedList<String> ();
for (String s : c) { ... }
```

- Bedingung

```
while (x > 0) { if (...) { x = ... } ... }
```

- Zustand (Iterator, hasNext, next)

## Zählschleifen

- Idee: vor Beginn steht Anzahl der Durchläufe fest.
- richtig realisiert ist das nur in Ada:

```
for p in 1 .. 10 loop ... end loop;
```

- Zähler `p` wird implizit deklariert
  - Zähler ist nur im Schleifenkörper sichtbar
  - Zähler ist im Schleifenkörper konstant
  - Zählerstand wird nur implizit durch Schleifensteuerung geändert
  - Ausdrücke für Bereichsgrenze wird nur einmal (zu Beginn) ausgewertet
- Vergleiche (beide Punkte) mit Java, C++, C

## Termination

- Satz: Jedes Programm aus: Zuweisungen, Blöcken, Verzweigungen, Zählschleifen hält für jede Eingabe.
- äquivalent zu Zählschleife: strukturelle Induktion (fold, primitive Rekursion)
- Satz: es gibt totale berechenbare (= allgemein rekursive) Funktionen, die nicht primitiv rekursiv sind.  
Beispiele: Interpreter für primitiv rekursive Programme, die Ackermann-Funktion
- softwaretechnisch ist statisch garantierte Termination wünschenswert. Z.B. erreicht durch Zählschleifen statt `while`, Rekursionsschema (fold) statt Rekursion

## Datengesteuerte Schleifen

- die Zählschleife ist schon ein *code smell* (Anzeichen für unzuweckmäßige Programmierung),
- der eigentliche *smell* ist die Verwendung von Zahlen (!)
- weil man (wegen verfrühter Optimierung) über Indizes spricht statt über die indizierten Werte

```
T [] a; for (int i = ...) { ... a[i] ... }
```

- Notation zur Vermeidung von Indizes bei Verarbeitung *aller* Elemente einer Datenstruktur

```
for (T x : a) { ... x ... }
```

mit Indizes, die gar nicht dastehen, kann man auch keine Indexfehler machen (z.B. off-by-one)

## Zustandgesteuerte Schleifen

- Iterator repräsentiert Strom von Daten  
(Stream = Liste mit bedarfsweiser Auswertung)

```
interface Iterator<T> {
 boolean hasNext(); T next ();
}
interface Iterable<T> {
 Iterator<T> iterator();
}
```

- Iterator-Objekt ist oft Index(Variable) mit Verweis auf zugrundeliegende Struktur. Das vermeidet Risiken wie: `int i; int j; int [] a; int [] b; .. a[j]`
- Iterator ist hier implizit (über den Wert einer Variablen, die gar nicht dasteht, muß man nicht nachdenken)

```
Iterable <T> c; for (T x : c) { ... }
```

## Implizite Iteratoren in C#

- durch diese Konstruktion wird ein Iterator angelegt:

```
using System.Collections.Generic;
public class it {
 public static IEnumerable<int> Data () { // <===
 yield return 3; yield return 1;
 yield return 4;
 }
 public static void Main () {
 foreach (int i in Data()) {
 System.Console.WriteLine (i); } } }
```

- der markierte Block ist eine Co-Routine, seine Ausführung ist mit der des Hauptprogramms verschränkt.
- Coroutinen in Simula (1967), siehe: Ole-Johan Dahl, C. A. R. Hoare: *Hierarchical Program Structures*, 1972 <https://dl.acm.org/doi/book/10.5555/1243380>

## Bedingungsgesteuerte Schleifen

- das ist die allgemeinste Form, ergibt (partielle) rekursive Funktionen, die terminieren nicht notwendig für alle Argumente.
- Steuerung
  - am Anfang: `while (Bedingung) Anweisung`
  - am Ende: `do Anweisung while (Bedingung)`
- Weitere Änderung des Ablaufes:
  - vorzeitiger Abbruch (`break`)
  - vorzeitige Wiederholung (`continue`)
  - beides auch nicht lokal

## Dynamische Semantik von Schleifen

- operationale Semantik durch Sprünge (autotool-Aufgabe)

```
while (B) A; ==>
 start : if (!B) goto end;
 A;
 goto start;
 end : skip;
```

- Ü: `do A while (B);`
- diese Programme sind semantisch äquivalent:

```
while (B) A; und if (B) { A; while (B) A }
```

- das definiert auch die Semantik (durch Transformation)
- Compiler machen das tatsächlich (loop unrolling)

## vorzeitiges Verlassen

- ... des Schleifenkörpers

```
while (B1) { A1; if (B2) continue; A2; }
```

- ... der Schleife

```
while (B1) { A1; if (B2) break; A2; }
```

- operationale Semantik: äquivalentes Goto-Programm

```
start: if (B1) goto next else goto end;
next : A1; if (B2) goto ... ; A2; goto start;
end : skip;
```

- äquivalentes Programm mit Standard-Schleife?

- für `continue`: einfach
- für `break`: nur mit Boolescher Hilfsvariablen

## Geschachtelte Schleifen

- manche Sprachen gestatten Markierungen (Labels) an Schleifen als Ziele für `break`, `continue`:

```
foo : for (int i = ...) {
 bar : for (int j = ...) {
 ... ; if (...) break foo; ... } }
}
```

- Ü: Syntax und Semantik mittels JLS erklären:

```
public static void main (String [] args) {
 System.out.println ("mozilla:open");
 https://haskell.org
 System.out.println ("mozilla:close"); }
}
```

## Sprünge

- bedingte, unbedingte (mit bekanntem Ziel)
  - Maschinensprachen, Assembler, Java-Bytecode
  - Fortran, Basic: if Bedingung then Zeilennummer
  - Fortran: dreifach-Verzweigung (arithmetic-if)
- “computed goto” (Zeilenr. des Sprungziels ausrechnen)
- zur Geschichte der Verzweigung in Programmiersprachen (mit vielen Original-Dokumenten)  
Eric Fischer: *if-then-else had to be invented*, !!Con West 2019 <https://github.com/ericfischer/if-then-else/blob/master/if-then-else.md> <http://bangbangcon.com/west/2019/speakers/>

## Sprünge und Schleifen

- Goto und While: softwaretechnisch (pragmatisch) sehr unterschiedlich, aber semantisch gleich ausdrucksstark
- Satz: zu jedem While-P. gibt es ein äquivalentes Goto-P.

- Satz: zu jedem Goto-P. gibt es ein äquivalentes While-P.

Beweis durch Kompilation:

übersetze 1: A1; 2: A2; .. 5: goto 7; zu

```
while (true) { switch (pc) {
 case 1 : A1 ; pc++ ; break; ...
 case 5 : pc = 7 ; break; ... } }
```

- das beweist: ... äquivalentes While-P. mit  $\leq 1$  Schleife
- softwaretechnisch nützt das gar nichts

### Schleifen und Unterprogramme

- Zu jedem While-P. gibt es ein äquivalentes P. ohne Schleifen: nur mit Verzweigungen und (rekursiven) Unterprogrammen
- Beweis-Idee: `while (B) A;` wird übersetzt in
 

```
void s () { if (B) { A; s (); } }
```
- Anwendung: C-Programme ohne Schlüsselwörter.  
(Wiederholung: wie entfernt man `if`?)
- Anwendung: International Obfuscated C Code Contest <https://www.ioccc.org/>

### Aufgaben zur Programm-Äquivalenz

- vereinfachtes Modell, damit Eigenschaften entscheidbar werden (sind die Programme  $P_1, P_2$  äquivalent?)
- Syntax: Programme
  - Aktionen,
  - Zustandsprädikate (in Tests)
  - Sequenz/Block, Verzweigung (if)
  - Sprünge: Label, goto,
  - Schleifen: while, break, continue
  - Boolesche Variablen und Operatoren
- Beispiel: `while (B && !C) { P; if (C) Q; }`

## Approximierte Spur-Semantik v. Programmen

- Semantik des Programms  $P$  ist Menge der Spuren von  $P$ .
  - *Spur* = eine Folge von Paaren von Zustand und Aktion,
  - ein *Zustand* ist eine Belegung der Prädikatsymbole,
  - jede Aktion zerstört alle Zustandsinformation.
- Satz: Diese Spursprachen (von goto- und while-Programmen) sind *effektiv regulär*.  
Beweis: Konstruktion über endlichen Automaten.
  - Zustandsmenge = Prädikatbelegungen  $\times$  Anweisungs-Nummer
  - Transitionen? (Beispiele)

Damit ist Spur-Äquivalenz von Programmen entscheidbar.— Beziehung zu tatsächlicher Äquivalenz?

## Hausaufgaben

### 1. Syntax If-Then-Else

- (Wdhlg) Ergänzen: das Problem des *dangling else* ist die Mehrdeutigkeit der Grammatik mit den Regeln ...
- (Wdhlg) Geben Sie ein Beispielprogramm  $P$  mit 2 Ableitungsbäumen bzgl. einer solchen Grammatik an.
- Suchen Sie die entsprechenden Regeln der Java-Grammatik,
- geben Sie den Ableitungsbaum für voriges  $P$  bzgl. dieser Grammatik an, begründen Sie, daß diese Grammatik eindeutig ist.
- Suchen Sie die entsprechenden Regeln in der Grammatik der Programmiersprache Ada,
- wie muß  $P$  geändert werden, damit es durch diese Grammatik erzeugt werden kann?

### 2. Kompilation für Mehrfachverzweigung

- Schreiben Sie ein Programm, das einen C-Programmtext dieser Form ausgibt

```

void p(int x) {
 switch (x) {
 case 0 : q0(); break;
 case 1 : q1(); break;
 ...
 case 999 : q999(); break; } }

```

Unterprogramme  $q_i$  nicht definieren, es geht nur um Kompilation (zu Objektfile, ohne Linking)

(b) Betrachten Sie den Assemblercode, der dafür von `gcc -O2 -S` erzeugt wird.

(c) Ändern Sie das Programm zu

```

case 0 : ...
case 100 :
...
case 99900 :

```

beobachten und erklären Sie (ggf. weiter Abstände ausprobieren)

3. ein (kurzes) (Gewinner-)Programm eines IOCCC vorführen und erläutern, bei dem Programmablaufsteuerung nicht offensichtlich ist

4. Adventskalenderaufgabe mit esoterischer Sprache

## 9 Unterprogramme

### Grundsätzliches

- Ein Unterprogramm ist ein Block mit einer Schnittstelle.  
(*benannt* oder *anonym* = Lambda-Ausdruck)
- Funktion: liefert Wert, Aufruf ist Ausdruck  
Prozedur: liefert keinen Wert, Aufruf ist Anweisung
- Schnittstelle beschreibt Datentransport zwischen Aufrufer und Unterprogramm.
- zu Schnittstelle gehören
  - Deklarat. d. formalen Parameter (Name, Typ, Modus)
  - bei Funktionen: Deklaration des Resultattyps

## Parameter-Übergabe (Semantik)

Datenaustausch zw. Aufrufer (caller) und Aufgerufenem (callee): über globalen Speicher

```
#include <errno.h>
extern int errno;
```

oder über Parameter.

Datentransport (entspr. Schlüsselwörtern in Ada)

- in: (Argumente) vom Aufrufer zum Aufgerufenen
- out: (Resultate) vom Aufgerufenen zum Aufrufer
- in out: in beide Richtungen

## Parameter-Übergabe (Implementierungen)

- pass-by-value (Wert)
- copy in/copy out (Wert)
- pass-by-reference (Verweis)  
d.h. der formale Parameter im Unterprogramm bezeichnet *die gleiche Speicherstelle* wie das Argument beim Aufrufer (Argument-Ausdruck muß lvalue besitzen)
- pass-by-name (textuelle Substitution)  
selten ... Algol68, CPP-Macros ... Vorsicht!

## Parameterübergabe

häufig benutzte Implementierungen:

- Pascal: by-value (default) oder by-reference (VAR)
- C: immer by-value (Verweise ggf. selbst herstellen)
- C++ by-value *oder* by-reference (durch &)  

```
void p(int & x) { x++; } int y = 3; p(y);
```
- Java: immer by-value  
(beachte implizite Zeiger bei Verweistypen)

- C#: by-value (beachte implizite Zeiger bei Verweistypen, `class`, jedoch nicht bei `struct`)  
oder by-reference (mit Schlüsselwort `ref`)
- Scala: by-value oder by-name (Scala Lang Spec 6.6)

### Call-by-value, call-by-reference (C#)

- by value:

```
static void u (int x) { x = x + 1; }
int y = 3 ; u (y);
Console.WriteLine(y); // 3
```

- by reference:

```
static void u (ref int x) { x = x + 1; }
int y = 3 ; u (ref y);
Console.WriteLine(y); // 4
```

### Call-by-name

formaler Parameter wird durch Argument-Ausdruck ersetzt.

Algol(68): Jensen's device

```
int sum (int i, int n; int f) {
 int s = 0;
 for (i=0; i<n; i++) { s += f; }
 return s;
}
int [10][10] a; int k; sum (k, 10, a[k][k]);
```

moderne Lösung

```
int sum (int n; Func<int,int> f) {
 ... { s += f (i); }
}
int [10][10] a; sum (10, (int k) => a[k][k]);
```

## Call-by-name (Macros)

```
#define thrice(x) 3*x // gefährlich
thrice (4+y) ==> 3*4+y
```

“the need for a preprocessor shows omissions in the language”

- fehlendes Modulsystem (Header-Includes)
- fehlende generische Polymorphie ( $\Rightarrow$  Templates in C+)

weitere Argumente:

- mangelndes Vertrauen in optimierende Compiler (inlining)
- bedingte Übersetzung

Ü: was kann der Präprozessor in C# und was nicht? Warum? (Wo ist der C#-Standard?)

<http://stackoverflow.com/questions/13467103>

## Call-by-name in Scala

Parameter-Typ ist  $\Rightarrow T$ , entspr. „eine Aktion, die ein T liefert“ (in Haskell: IO T)  
call-by-name

```
def F(b:Boolean, x: =>Int):Int =
 { if (b) x*x else 0 }
F(false, {print ("foo "); 3})
// res5: Int = 0
F(true, {print ("foo "); 3})
// foo foo res6: Int = 9
```

Man benötigt call-by-name zur Definition von Abstraktionen über den Programmablauf.

Übung: If, While als Scala-Unterprogramm

## Bedarfsauswertung

- andere Namen: (call-by-need, lazy evaluation)
- Definition: das Argument wird bei seiner ersten Benutzung ausgewertet
- wenn es nicht benutzt wird, dann nicht ausgewertet;  
wenn mehrfach benutzt, dann nur einmal ausgewertet
- das ist der Standard-Auswertungsmodus in Haskell:  
alle Funktionen und Konstruktoren sind *lazy*  
da es keine Nebenwirkungen gibt, bemerkt man das zunächst nicht ...  
... und kann es ausnutzen beim Rechnen mit unendlichen Datenstrukturen (Streams)

## Beispiele f. Bedarfsauswertung (Haskell)

- ```
[ error "foo" , 42 ] !! 0
[ error "foo" , 42 ] !! 1
length [ error "foo" , 42 ]
let xs = "bar" : xs
take 5 xs
```

- Fibonacci-Folge

```
fib :: [ Integer ]
fib = 0 : 1 : zipWith (+) fib ( tail fib )
```

- Primzahlen (Sieb des Eratosthenes)

- Papier-Falt-Folge

```
let merge (x:xs) ys = x : merge ys xs
let updown = 0 : 1 : updown
let paper = merge updown paper
take 15 paper
```

vgl. <https://www.imn.htwk-leipzig.de/~waldmann/etc/stream/>

Beispiele f. Bedarfsauswertung (Scala)

Bedarfsauswertung für eine lokale Konstante (Schlüsselwort `lazy`)

```
def F(b:Boolean,x: =>Int):Int =
  { lazy val y = x; if (b) y*y else 0 }
F(true,{print ("foo "); 3})
//   foo res8: Int = 9
F(false,{print ("foo "); 3})
//   res9: Int = 0
```

Argumente/Parameter

- in der Deklaration benutzte Namen heißen (formale) *Parameter*,
- bei Aufruf benutzte Ausdrücke heißen *Argumente*
(... nicht: aktuelle Parameter, denn engl. *actual* = dt. tatsächlich)

Designfragen bei Parameterzuordnung:

- über Position oder Namen? gemischt?
- defaults für fehlende Argumente?
- beliebig lange Argumentlisten?

Positionelle/benannte Argumente

Üblich ist Zuordnung über Position

```
void p (int height, String name) { ... }  
p (8, "foo");
```

in Ada: Zuordnung über Namen möglich

```
procedure Paint (height : Float; width : Float);  
Paint (width => 30, height => 40);
```

nach erstem benanntem Argument keine positionellen mehr erlaubt

code smell: lange Parameterliste,
refactoring: Parameterobjekt einführen
allerdings fehlt (in Java) benannte Notation für Record-Konstanten.

Default-Werte

C++:

```
void p (int x, int y, int z = 8);  
p (3, 4, 5); p (3, 4);
```

Default-Parameter müssen in Deklaration am Ende der Liste stehen

Ada:

```
procedure P  
  (X : Integer; Y : Integer := 8; Z : Integer);  
P (4, Z => 7);
```

Beim Aufruf nach weggelassenem Argument nur noch benannte Notation

Variable Argumentanzahl (C)

wieso geht das eigentlich:

```
#include <stdio.h>
char * fmt = really_complicated();
printf (fmt, x, y, z);
```

Anzahl und Typ der weiteren Argumente werden überhaupt nicht geprüft:

```
extern int printf
    (__const char *__restrict __format, ...);
```

Variable Argumentanzahl (Java)

```
static void check (String x, int ... ys) {
    for (int y : ys) { System.out.println (y); }
}
```

```
check ("foo",1,2); check ("bar",1,2,3,4);
```

letzter formaler Parameter kann für beliebig viele des gleichen Typs stehen.

tatsächlich gilt `int [] ys`,
das ergibt leider Probleme bei generischen Typen

Aufgaben zu Parameter-Modi

1. Semantik dieses Ada-Programm erklären (verschiedene GCC/GNAT-Versionen) unter Bezug auf Sprachstandard (2012, vgl. mit früheren) und *Rationale*.

```
with Ada.Text_IO; use Ada.Text_IO;
procedure Check is
    procedure Sub (X: in out Integer;
                  Y: in out Integer;
                  Z: in out Integer) is
    begin
        Y := 8; Z := X;
    end;
    Foo: Integer := 9;   Bar: Integer := 7;
begin
```

```

    Sub (Foo, Foo, Bar);
    Put_Line (Integer' Image(Foo));
    Put_Line (Integer' Image(Bar));
end Check;

```

(in Datei Check.adb schreiben, kompilieren mit gnatmake Check.adb)

Vergleichen mit diesem C++-Programm:

```

#include <iostream>

void sub (int & x, int & y, int & z) {
    y = 8;
    z = x;
}

int main () {
    int foo = 9;
    int bar = 7;

    sub (foo, foo, bar);
    std::cout << foo << std::endl;
    std::cout << bar << std::endl;
}

```

2. Call by value, call by reference

```

class C { public int foo; }
class M { public static void u (C x)
    { x.foo=4; x=new C{foo=5}; } }

C y = new C {foo=3}
C z = y
u (y)
y.foo
z.foo

```

- Kompilieren/ausführen (mit csharp CLI), beobachten, erklären. Diagramm zeichnen, das die Speicherbelegung verdeutlicht.

- Ersetzen Sie `class C` durch `struct C`. Kompilieren, ...
- Ersetzen Sie `void u (C x)` durch `void u (ref C x)`. Welche weitere Änderung ist erforderlich? Kompilieren, ...
- `class C` und `u (ref C x)`

3. call by name, call by reference:

Wie kann man diese beiden Unterprogramme aus Sicht des Aufrufers semantisch voneinander unterscheiden:

- Funktion (C++): (call by reference)

```
void swap (int & x, int & y)
    { int h = x; x = y; y = h; }
```

- Makro (C): (call by name)

```
#define swap(x, y) \
    { int h = x; x = y; y = h; }
```

Geben Sie einen Ausdruck E an, in dem ein Name `swap` benutzt wird, so daß für beide Definitionen von `swap` gilt:

- E ist syntaktisch korrekt,
- E ist statisch korrekt,
- dynamische Semantiken von E sind unterschiedlich

Also *nicht* einfach so:

```
int a = 3; int b = 4; swap (a,b);
```

4. Simulation von call-by-name durch Unterprogramme als Argumente:

- (a) Die Fakultäts-Funktion in ECMA-Script

```
function f(x) { return x==0 ? 1 : x * f(x-1) }
f(4)
==> 24
```

- (b) Die Verzweigung als Funktion

```
function ite(b, j, n) { return b ? j : n }
ite(false, 2, 3)
==> 3
```

(c) Ersetzen Sie `?:` in `f` durch `ite`, werten Sie `f(4)` aus, erklären Sie Ihre Beobachtung.

(d) Simulation von call-by-name durch Unterprogramme als Argumente:

```
function ite(b, j, n) { return b ? j() : n() }
```

wie muß `ite(false, 2, 3)` jetzt aussehen?

(e) passen Sie die Def. von `f` an und testen Sie

10 Weiteres zu Unterprogrammen

Lokale Unterprogramme

- Unterprogramme sind wichtiges Mittel zur Abstraktion, das möchte man überall einsetzen
- also sind auch lokale Unterprogramme wünschenswert
(die Konzepte *Block* (mit lokalen Deklarationen) und *Unterprogramm* sollen *orthogonal* sein)

```
int f (int x) {
  int g (int y) { return y + 1; }
  return g (g (x));
}
```

UP und Sichtbarkeit von Namen

```
{ const x = 3;
  function step(y) { return x + y; }
  for (const z of [ 1, 2, 4 ]) {
    console.log(step(z+1)); } }
```

- was ist die Ausgabe dieses Programms?
- was ändert sich bei Umbenennung von `z` zu `x`?

- Antwort: nichts! — der Funktionskörper ($x+y$) wird in seiner Definitionsumgebung ausgewertet, nicht in seiner Aufruf-Umgebung.

vgl. Spezifikation: <https://tc39.github.io/ecma262/#sec-lexical-environments>,
<https://www.ecma-international.org/ecma-262/7.0/>

Frames, Ketten, Indizes

Während ein Unterprogramm rechnet, stehen seine lokalen Daten in einem Aktivationsverbund (Frame).

- Jeder Frame hat zwei Vorgänger:
 - dynamischer Vorgänger:
(Frame des *aufrufenden* UP) benutzt zum Rückkehren
 - statischer V. (Frame des *textuell umgebenden* UP)
benutzt zum Zugriff auf “fremde” lokale Variablen
- Jeder Variablenzugriff hat *Index-Paar* (i, j) :
im i -ten statischen Vorgänger der Eintrag Nr. j ,
lokale Variablen des aktuellen UP: Index $(0, j)$

Indizes werden *statisch* bestimmt, Frames zur Laufzeit.
 Der Zugriff auf den indizierten Wert ist *keine Suche!*

Lokale Unterprogramme: Beispiel

```
with Ada.Text_IO; use Ada.Text_IO;
procedure Nested is
  function F (X: Integer; Y: Integer)
  return Integer is
    function G (Y: Integer) return Integer is
    begin
      if (Y > 0) then return 1 + G(Y-1);
      else return X; end if;
    end G;
  begin return G (Y); end F;
begin
  Put_Line (Integer'Image (F(3,2)));
end Nested;
```

Globale Unterprogramme

Entwurfs-Entscheidung für C (1972):

- keine lokalen UP, jedes UP ist global

Auswirkungen:

- leichte Implementierung:
 - dyn. Vorgänger = der vorige Frame (auf dem Stack)
 - statischer Vorgänger: gibt es nicht
- softwaretechnische Nachteile:
 - globale Abstraktionen machen Programm unübersichtlich (vgl.: globale Variablen).
 - Gegen-Argument: Nachnutzbarkeit, Testbarkeit

Lokale Unterprogramme

- Lambda-Kalkül 1936 $(\lambda f.f(f0))(\lambda x.x + 1)$,
- LISP 1958, Algol 60, Pascal 1970, ML 1973, Ada 1983,
- Haskell 1990: $(\backslash f \rightarrow f (f 0)) (\backslash x \rightarrow x + 1)$
- JavaScript (ES6, 2015) $(f=>f(f(0))) (x=>x+1)$

```
(function(f){return f(f(0))})  
  (function(x){return x+1})
```

- C# (3.0, 2007)

```
int x = 3; Func<int,int> f = y => x + y;  
Console.WriteLine (f(4));
```

- Java (8, 2014)

```
int x = 3; Function<Integer,Integer> f = y -> x + y;  
System.out.println (f.apply(4));
```

Unterprogramme als Argumente

```
static int d ( Func<int,int> g ) {
    return g(g(1));
}
static int p (int x) {
    Func<int,int> f = y => x + y;
    return d (f);
}
```

Betrachte Aufruf $p(3)$.

Das innere Unterprogramm f muß auf den p -Frame zugreifen, um den richtigen Wert des x zu finden.

Dazu *Closure* konstruieren: f mit statischem Vorgänger.

Wenn Unterprogramme als Argumente übergeben werden, steht der statische Vorgänger im Stack.

(ansonsten muß man den Vorgänger-Frame auf andere Weise retten, siehe später)

Unterprogramme als Resultate

```
static int x = 3;
static Func<int,int> s (int y) {
    return z => x + y + z;
}
static void Main () {
    Func<int,int> p = s(4);
    Console.WriteLine (p(3));
}
```

Wenn die von $s(4)$ konstruierte Funktion p aufgerufen wird, dann wird der s -Frame benötigt, steht aber nicht mehr im Stack.

⇒ Die (Frames in den) Closures müssen im Heap verwaltet werden.

Lokale anonyme Unterprogramme

- ```
int [] x = { 1,0,0,1,0 };
Console.WriteLine
 (x.Aggregate (0, (a, b) => 2*a + b));
http://code.msdn.microsoft.com/LINQ-Aggregate-Operators-c51b3869
```
- ```
foldl ( \ a b -> 2*a + b) 0 [1,0,0,1,0]
Haskell (http://haskell.org/)
```

historische Schreibweise: $\lambda ab.2a + b$
(Alonzo Church: The Calculi of Lambda Conversion, 1941)
vgl. Henk Barendregt: The Impact of the Lambda Calculus, 1997, <ftp://ftp.cs.ru.nl/pub/CompMath.Found/church.ps>

Lokale Klassen (Java)

- static nested class: dient lediglich zur Gruppierung

```
class C { static class D { .. } .. }
```

- nested inner class:

```
class C { class D { .. } .. }
```

jedes D-Objekt hat einen Verweis auf ein C-Objekt (\approx statische Kette) (bezeichnet durch `C.this`)

- local inner class: (Zugriff auf lokale Variablen in *m* nur, wenn diese final sind. Warum?)

```
class C { void m () { class D { .. } .. } }
```

Lokale Funktionen in Java 8

```
interface Function<T,R> { R apply(T t); }
```

bisher (Java ≤ 7):

```
Function<Integer,Integer> f =  
    new Function<Integer,Integer> () {  
        public Integer apply (Integer x) {  
            return x*x;  
        }  
    } ;
```

```
System.out.println (f.apply(4));
```

jetzt (Java 8): verkürzte Notation (Lambda-Ausdruck) für Implementierung *funktionaler Interfaces*

```
Function<Integer,Integer> g = x -> x*x;  
System.out.println (g.apply(4));
```

Anwendung u.a. in `java.util.stream.Stream<T>`

Unterprogramme/Zusammenfassung

in prozeduralen Sprachen:

- falls alle UP global: dynamische Kette reicht
- lokale UP: benötigt auch statische Kette
- lokale UP as Daten: benötigt Closures
= (Code, statischer Link)
- UP als Argumente: Closures auf Stack
- UP als Resultate: Closures im Heap

in objektorientierten Sprachen: ähnliche Überlegungen bei lokalen (inner, nested) Klassen.

Hausaufgaben

1. Assembler-Code für Programm von „Lokale UP: Beispiel“ mit `gcc -c -O0 -S nested.adb`,
 - welche Variablen-Benutzung hat Index (i, j) mit $i > 0$,
 - wo steht das im Assemblercode?
 - vergleiche Assemblercode des Hauptprogramms bei `-O0 / -O3`
2. Lokale UP (Lambda-Ausdrücke) in C++:

```
#include <iostream>
#include <functional>
using namespace std;
int x = 3;
function<int(int)> s (int y) {
    return [] (int z) { return x+y+z; };
}
int main () {
    auto p = s(1);
    auto q = s(5);
    cout << p(2) << endl;
}
```

- Dieses Programm ist statisch falsch, warum?
Ersetzen Sie [] durch [=]
- Ersetzen Sie [] durch [&], begründen Sie das beobachtete Verhalten mit Hilfe des Standards (Dokument N4800) <http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2019/> (oder aktuellere Version)

3. Das *most recent*-Problem erklären.

van den Hove d'Ertsenryck: *Dissolving a half century old problem about the implementation of procedures*, 2017 <https://ir.cwi.nl/pub/26757>

Beispiele in der autotool-Aufgabe vorführen (soweit möglich)

4. Aussagen über Graphen mit Knotenmenge = Frames einer Programmausführung, Kanten \rightarrow_{dyn} , $\rightarrow_{\text{stat}}$.

- \rightarrow_{dyn} ist ein Baum
- $\rightarrow_{\text{stat}}$ ist ein Baum
- sind beliebige Kombinationen von Bäumen möglich? Nein, \rightarrow_{dyn} und $\rightarrow_{\text{stat}}$ besitzen eine gemeinsame topologische Ordnung. Woher kommt diese?
- sind alle Kombinationen mit gemeinsamer topologischer Ordnung möglich?

Zur autotool-Aufgabe zu Frames: siehe auch <https://gitlab.imn.htwk-leipzig.de/autotool/all10/issues/124>

11 Dynamische Polymorphie

Übersicht

poly-morph = viel-gestaltig; ein Bezeichner (z. B. Unterprogramm-Name) mit mehreren Bedeutungen

Arten der Polymorphie:

- statische P.
(Bedeutung wird zur Übersetzungszeit festgelegt):
 - ad-hoc: Überladen von Bezeichnern
 - generisch: Bezeichner mit Typ-Parametern
- dynamische P. (Bedeutung wird zur Laufzeit festgelegt):
 - Implementieren (Überschreiben) von Methoden, Auswahl der Impl. anhand des dynamischen Typs

Objekte, Methoden

Motivation: Objekt = Daten + Verhalten.

Einfachste Implementierung:

- Objekt ist Record,
- einige Komponenten sind Unterprogramme.

```
typedef struct {
    int x; int y; // Daten
    void (*print) (FILE *fp); // Verhalten
} point;
point *p; ... ; (*(p->print))(stdout);
```

Anwendung: Datei-Objekte in UNIX (seit 1970)

(Merksatz 1: all the world is a file) (Merksatz 2: those who do not know UNIX are doomed to re-invent it, poorly)

Objektbasierte Sprachen (JavaScript)

(d. h. objektorientiert, aber ohne Klassen)

Objekte, Attribute, Methoden:

```
var o = { a : 3,
  m : function (x) { return x + this.a; } };
```

Vererbung zwischen Objekten:

```
var p = { __proto__ : o };
```

Attribut (/Methode) im Objekt nicht gefunden \Rightarrow weitersuchen im Prototyp \Rightarrow ... Prototyp des Prototyps ...

Übung: Überschreiben

```
p.m = function (x) { return x + 2*this.a }
var q = { __proto__ : p }
q.a = 4
q.m(5)
```

Klassenbasierte Sprachen

gemeinsame Datenform und Verhalten von Objekten

```
typedef struct { int (*method[5]) (); } cls;
typedef struct {
    cls * c;
} obj;
obj *o; ... (*(o->c->method[3])) ();
```

allgemein: Klasse:

- Deklaration von Daten (Attributen)
- Deklaration und Implementierung von Methoden

Objekt:

- tatsächliche Daten (Attribute)
- Verweis auf Klasse (Methodentabelle)

this

Motivation: Methode erfährt, für welches Argument sie gerufen wurde

```
typedef struct { int (*method[5]) (obj *o);
} cls;
typedef struct {
    int data [3]; // Daten des Objekts
    cls *c; // Zeiger auf Klasse
} obj;
obj *o; ... (*(o->c->method[3])) (o);
int sum (obj *this) {
    return this->data[0] + this->data[1]; }
```

jede Methode bekommt *this* als erstes Argument
(in Java, C# geschieht das implizit)

Klassen in ECMA-Script

- syntaktische Hilfen zur Notation der objekt(prototyp)-basierten Vererbung, seit Version 6 (2015)

- ```
class C {
 constructor(x) { this.x=x }
 m (y) { return this.x + y } }
let p = new C(8)
p.m(3)
```
- **Definition** siehe <https://www.ecma-international.org/ecma-262/7.0/#sec-class-definitions>

## Vererbung

Def: Klasse *D* ist *abgeleitet* von Klasse *C*:

- *D* kann Menge der Attribute- und Methodendeklarationen von *C* erweitern (aber nicht verkleinern oder ändern)
- *D* kann Implementierungen von in *C* deklarierten Methoden übernehmen oder eigene festlegen (überschreiben).

Anwendung: dynamische Polymorphie

- Wo ein Objekt der Basisklasse erwartet wird (der *statische Typ* eines Bezeichners ist *C*),
- kann ein Objekt einer abgeleiteten Klasse (*D*) benutzt werden (der *dynamische Typ* des Wertes ist *D*).

## Dynamische Polymorphie (Beispiel)

```
class C {
 int x = 2; int p () { return this.x + 3; }
}
C x = new C() ; int y = x.p ();
```

Überschreiben:

```
class E extends C {
 int p () { return this.x + 4; }
}
C x = // statischer Typ: C
 new E() ; // dynamischer Typ: E
int y = x.p ();
```

## Vererbung bricht Kapselung

- ```
class C { void p () { ... q(); ... };  
        void q () { .. }; }
```

- Jetzt wird `q` überschrieben (evtl. auch unabsichtlich—in Java), dadurch ändert sich das Verhalten von `p`.

```
class D extends C { void q () { ... } }
```

- Korrektheit von `D` abhängig von *Implementierung* von `C`
- ⇒ object-orientation is, by its very nature, anti-modular

Bob Harper, 2011: <https://web.archive.org/web/20140819133753/http://existentialtype.wordpress.com/2011/03/15/teaching-fp-to-freshmen/>

Einordnung Objektorientierung

- OO: *der* Hype der 80er (vgl. XML, Container, Cloud, Edge, Blockchain, Elektrotretroller, as a service, KI)
- *nützlich*:
 - `class C` benutzerdefinierte, anwendungsspezif. Typen
 - `class C { A x; B y; E m() { .. } }`
Gruppierung von Daten und UP zu ihrer Verarbeitung
Simulation von Funktionen als Daten
 - `interface I; class C implements I; I x = new E();` Trennung:
Schnittstelle (abstrakter Datentyp, Signatur), Implementierung (konkreter Datentyp, Algebra)
- *schädlich*: Zustandsänder.g, Implementierungs-Vererb.g
- hat Praxis und Lehre der Programmierung nachhaltig beschädigt und das ist noch nicht ausgestanden

Übung Dynamische Polymorphie

1. (HA für WS20) Beispiel auf Folie „Objektbasierte Sprachen (JS)“ ausprobieren, Beobachtungen erklären (Speicherbelegung grafisch darstellen), erweitern

2. zu Folie „Vererbung bricht Kapselung:

vgl. Joshua Bloch: *Effective Java* (Pearson 2018)

Item 19: „Design and document for inheritance or else prohibit it.“

Diskutieren Sie die Einhaltung dieser Regel am Beispiel [https://docs.oracle.com/en/java/javase/11/docs/api/java.base/java/util/AbstractCollection.html#retainAll\(java.util.Collection\)](https://docs.oracle.com/en/java/javase/11/docs/api/java.base/java/util/AbstractCollection.html#retainAll(java.util.Collection))

Statische Polymorphie: Ad-Hoc-Polymorphie

- ein Bezeichner ist *überladen*, wenn er mehrere (gleichzeitig sichtbare) Deklarationen hat
- bei jeder Benutzung des Bezeichners wird die Überladung dadurch *aufgelöst*, daß die Deklaration mit dem jeweils (ad-hoc) passenden Typ ausgewählt wird

Beispiel: Überladung im Argumenttyp:

```
static void p (int x, int y) { ... }
static void p (int x, String y) { ... }
p (3, 4); p (3, "foo");
```

keine Überladung nur in Resultattyp, denn...

```
static int f (boolean b) { ... }
static String f (boolean b) { ... }
```

Typhierarchie als Halbordnung

- extends/implements definiert Halbordnung auf Typen, Bsp.
class C; class D extends C; class E extends C definiert Relation auf $T = \{C, D, E\}$
 $(\leq) = \{(C, C), (D, C), (D, D), (E, C), (E, E)\}$
- dadurch entsteht Halbordnung auf Methoden-Signaturen (Tupel der Argument-Typen, ohne Resultat-Typ)

Bsp: Relation \leq^2 auf T^2 :

$$(t_1, t_2) \leq^2 (t'_1, t'_2) : \iff t_1 \leq t'_1 \wedge t_2 \leq t'_2$$

es gilt $(D, D) \leq^2 (C, C)$; $(D, D) \leq^2 (C, D)$; $(C, D) \leq^2 (C, C)$; $(E, C) \leq^2 (C, C)$.

Ad-Hoc-Polymorphie und Typhierarchie

Auflösung von `p (new D(), new D())` bzgl.

```
static void p (C x, D y);
static void p (C x, C y);
static void p (E x, C y);
```

- bestimme die Menge P der zum Aufruf *passenden* Methoden
(für diese gilt: statischer Typ der Argumente \leq^n statischer Typ der formalen Parameter)
- bestimme die Menge M der minimalen Elemente von P
(Def: m ist minimal falls $\neg \exists p \in P : p < m$)
- M muß eine Einermenge sein, sonst ist Überladung nicht auflösbar

Überschreiben und Überladen

- Überschreiben:
zwei Klassen, Methoden mit übereinstimmendem Namen und Typ
- Überladen:
 ≥ 1 Klasse, gleichnamige M. mit unterschiedl. Typen
- C++: Methoden, die man überschreiben darf, `virtual` deklarieren
- C#: Überschreiben durch `override` anzeigen,
- Java: alle Methoden sind `virtual`, deswegen ist Überschreiben von Überladen schlecht zu unterscheiden:
ist Quelle von Programmierfehlern
- Java-IDEs unterstützen Annotation `@overrides`

Equals richtig implementieren

```
class C {
    final int x; final int y;
    C (int x, int y) { this.x = x; this.y = y; }
    int hashCode () { return this.x + 31 * this.y; }
}
```

nicht so:

```
public boolean equals (C that) {
    return this.x == that.x && this.y == that.y;
}
```

Equals richtig implementieren (II)

...sondern so:

```
public boolean equals (Object o) {
    if (! (o instanceof C)) return false;
    C that = (C) o;
    return this.x == that.x && this.y == that.y;
}
```

Die Methode `boolean equals(Object o)` wird aus `HashSet` aufgerufen.

Sie muß deswegen *überschrieben* werden.

Das `boolean equals (C that)` hat den Methodenamen nur *überladen*.

Statische Attribute und Methoden

- für diese findet *kein* dynamischer Dispatch statt.

```
class C {static int f(){return 0;}}
class D extends C {static int f(){return 1;}}
C x = new D()
x.f()
```

- Damit das klar ist, wird dieser Schreibweise *aller* Methodenaufrufe empfohlen:
 - dynamisch: immer mit Objektnamen qualifiziert, auch wenn dieser `this` lautet,
 - statisch: immer mit Klassennamen qualifiziert (niemals mit Objektnamen)

Übung Ad-Hoc-Polymorphie

- (HA für WS20) Gegeben sind diese Klassen und Methoden eines Java-Programmes:

```
class D extends B; class B extends A; class A;
                    class C extends A;
static void p (B x, C y);
static void p (A x, D y);
static void p (B x, A y);
```

Beschreiben Sie, wie die Überladung in `p (new D(), new C())` aufgelöst wird.

- Wo und wie ist das Verfahren zur Auflösung der Ad-Hoc-Polymorphie im Java-Standard beschrieben?

12 Polymorphie

Übersicht

poly-morph = viel-gestaltig; ein Bezeichner (z. B. Unterprogramm-Name) mit mehreren Bedeutungen

Arten der Polymorphie:

- statische P.
(Bedeutung wird zur Übersetzungszeit festgelegt):
 - ad-hoc: Überladen von Bezeichnern
 - generisch: Bezeichner mit Typ-Parametern
- dynamische P. (Bedeutung wird zur Laufzeit festgelegt):
 - Implementieren (Überschreiben) von Methoden, Auswahl der Impl. anhand des dynamischen Typs

Generische Polymorphie (Typ-Argumente)

- Klassen und Methoden können Typ-Parameter erhalten.
- innerhalb der Implementierung der Klasse/Methode wird der formale Typ-Parameter als (unbekannter) Typ behandelt
- bei der Benutzung der Klasse/Methode müssen alle Typ-Argumente angegeben werden
(oder der Compiler inferiert diese,
R. Hindley (1969) *The Principal Type-Scheme of an Object in Combinatory Logic*
Transact. AMS 146:29-60.
- Motivation/Anwendung: Flexibilität, Sicherheit, Effizienz:
separate Kompilation (auch von generischen Methoden/Klassen) mit statischer Typprüfung

Bsp: Generische Methode in C#

```
class C {  
    static T id<T> (T x) { return x; }  
}
```

beachte Position(en) von

- *Deklaration* des Typparameters
- *Benutzungen* des Typparameters

```
string foo = C.id<string> ("foo");  
int bar = C.id<int> (42);
```

- *Instanziierung* des Typparameters

Bsp: Generische Klasse in Java

```
class Pair<A,B> {  
    final A first; final B second;  
    Pair(A a, B b)  
        { this.first = a; this.second = b; }  
}  
Pair<String,Integer> p =  
    new Pair<String,Integer>("foo", 42);  
int x = p.second + 3;
```

vor allem für Container-Typen (Liste, Menge, Keller, Schlange, Baum, ...)

Bsp: Generische Methode in Java

- *Deklaration* des Typparameters
- *Benutzungen* des Typparameters

```
class C {  
    static <A,B> Pair<B,A> swap (Pair<A,B> p) {  
        return new Pair<B,A>(p.second, p.first); } } }
```

- *Benutzungen* des Typparameters

```
Pair<String,Integer> p =
    new Pair<String,Integer>("foo", 42);
Pair<Integer,String> q =
    C.<String,Integer>swap(p);
```

Typargumente können auch *inferiert* werden:

```
Pair<Integer,String> q = C.swap(p);
```

Generische Fkt. höherer Ordg.

- Ziele:
 - Flexibilität (nachnutzbarer Code)
 - statische Typsicherheit
 - Effizienz (Laufzeit)
- wichtige Anwendung: Abstraktionen über den Programmablauf, z.B. für parallele Ausführung, Bsp:

```
public static
    TAccumulate Aggregate<TSource, TAccumulate> (
        this ParallelQuery<TSource> source,
        TAccumulate seed,
        Func<TAccumulate,TSource,TAccumulate> func )
```

Bsp. Generische Fkt. höherer Ordg. (I)

Sortieren mit Vergleichsfunktion als Parameter

```
using System; class Bubble {
    static void Sort<T>
        (Func<T,T,bool> Less, T [] a) { ...
        if (Less (a[j+1],a[j])) { ... } }
    public static void Main (string [] argv) {
        int [] a = { 4,1,2,3 };
        Sort<int> ((int x, int y) => x <= y, a);
        foreach (var x in a) Console.Write (x);
    } }
```

Ü: (allgemeinster) Typ und Implementierung einer Funktion Flip, die den Vergleich umkehrt: Sort<int> (Flip((x,y)=> x <= y), a)

Bsp. Generische Fkt. höherer Ordg. (II)

„bulk operations“ auf Collections, z.B.

- Bibliothek <https://hackage.haskell.org/package/containers/docs/Data-Map-Strict.html>
- Beispiel:

```
intersectionWith
  :: Ord k
 => (a -> b -> c)
 -> Map k a -> Map k b -> Map k c
```

- ist effizienter als Iteration über alle Elemente eines Arguments

Übung Typparameter

1. Sortieren mit Vergleichsfunktion als Parameter

(git clone <https://gitlab.imn.htwk-leipzig.de/waldmann/pps-ws18.git>)

- (a) Flip implementieren.
 - (b) Welches ist der allgemeinste Typ von Flip?
2. (HA für WS20) bulk operations auf Collections.

- (a) Bestimmen Sie den Typ von `Data.Map.unionWith` (API-Dokumentation oder ghci)
warum hat dieser weniger Typparameter als `intersectionWith`?
- (b) einfache Messungen mit ghci. Nach jeder Deklaration/Ausdruck anzeigen Kosten diskutieren

```
:set +s
import qualified Data.Set as S
a = S.fromList [1 :: Int .. 10^6 ]
length a
length a
b = S.map (+ 10^6) a
length b
S.intersection a b -- bulk operation
S.filter (\ x -> S.member x a) b -- naive elementweise Implementi
```

(c) warum ist die bulk operation hierfür langsamer?

```
c = S.map (* 2) a ; d = S.map succ c
```

und trotzdem noch schneller als elementweise?

(Nur die Kosten der Operation messen, nicht die der Konstruktion oder der Ausgabe.)

Ergänzung: `S.Set Int` ist unzweckmäßig, denn `Data.IntSet.Set` ist effizienter!

13 Mehr zu Polymorphie

Vererbung und generische Polym.

- mit Sprachkonzepte *Vererbung* ist Erweiterung des Sprachkonzeptes *Generizität* wünschenswert:
- beim Definition der Passung von parametrischen Typen sollte die Vererbungsrelation \leq auf Typen berücksichtigt werden.
- Ansatz: wenn $E \leq C$, dann auch $List<E> \leq List<C>$
- ist *nicht* typsicher, siehe folgendes Beispiel
- Modifikation: ko- und kontravariante Typparameter

Generics und Subtypen

Warum geht das nicht:

```
class C { }  
  
class E extends C { void m () { } }  
  
List<E> x = new LinkedList<E>();  
  
List<C> y = x; // Typfehler
```

Antwort: wenn das erlaubt wäre, dann:

variante Typ-Argumente (C#)

Kontravarianz (in P), Kovarianz (out P)

```
interface I<in P> { // Typ-Arg. ist kontravariant
    // P get (); kovariante Benutzung (verboten)
    void set (P x); // kontravariante Benutzung
}
class K<P> : I<P> { public void set (P x) {} }
class C {} class E : C {void m(){} } // E <= C
I<C> x = new K<C>();
I<E> y = x; // erlaubt, I<C> <= I<E>
```

- kontravariant: $E \leq C \Rightarrow I(E) \geq I(C)$
- kovariant: $E \leq C \Rightarrow I(E) \leq I(C)$
- invariant: $E \neq C \Rightarrow I(E) \not\leq I(C)$

Obere Schranken für Typparameter

- Java: `class<T extends S> { ... }`,
C#: `class <T> where T : S { ... }`
als Argument ist jeder Typ T erlaubt, der S implementiert

```
interface Comparable<T>
    { int compareTo(T x); }
static <T extends Comparable<T>>
    T max (Collection<T> c) { .. }
```

Untere Schranken für Typparameter

- Java: `<S super T>`
Als Argument ist jeder Typ S erlaubt, der Obertyp von T ist.

```
static <T> int binarySearch
    (List<? extends T> list, T key,
    Comparator<? super T> c)
```

Vergleich: Varianz und Schranken

Unterscheidung:

- Durch *Schranken* für Typ-Argumente
wird bei der Instantiierung des polymorphen Bezeichners (Typ, Methode)
die Wahl der Typargumente eingeschränkt.
- Durch *Varianz* für Typ-Argumente
wird die Zuweisungskompatibilität des instantiierten Typs erweitert (Sicht „von außen“)
und die Benutzung des Typ-Parameters eingeschränkt (Sicht „von innen“)

Generics und Arrays (in Java)

- das gibt keinen Typfehler:

```
class C { }  
class E extends C { void m () { } }  
E [] x = { new E (), new E () }; C [] y = x;  
y [0] = new C (); x [0].m();
```

- warum ist die Typprüfung für Arrays schwächer als für Collections? Historische Gründe. Das sollte gehen:

```
void fill (Object[] a, Object x) { .. }  
String [] a = new String [3];  
fill (a, "foo");
```

Übung Polymorphie

- (siehe Folie „untere Schranken...“)
binarySearch aufrufen (Java), so daß beide ? von T verschieden sind
- (siehe Folie „variante Typ-Parameter...“)
Implementieren Sie set und get in $K\langle P \rangle$, ergänzen Sie das Hauptprogramm so, daß schließlich eine Methode m() eines C-Objektes aufgerufen würde — was jedoch durch statische Typ-Prüfung verhindert wird.
- Wildcards (?) und Capture Conversion in JLS

14 Ergänzungen

Statisch typisiert ⇒ sicher und effizient

- Programmtext zeigt *Absicht* des Programmierers
- dazu gehören *Invarianten* von *Daten*, formuliert mittels *Typen* (`foo : Book`)
... alternative Formulierung: *Namen* (`fooBook`, *Kommentar* `foo // Book`)
- nur durch statische Typisierung kann man Absichten/ Annahmen maschinell prüfen u. Einhaltung erzwingen
- Yaron Minsky: *make illegal states un-representable*
- sichere und effiziente Ausführung, Wiederverwendbarkeit

Statische Typisierung: für und wider

Für statische Typisierung spricht vieles.

Es funktioniert auch seit Jahrzehnten (Algol 1960, ML 1970, C++ 1980, Java 1990 usw.)

Was spricht dagegen?

- Typsystem ist ausdruckschwach:
(Bsp: keine polymorphen Container in C)
Programmierer kann Absicht nicht ausdrücken
- Typsystem ist ausdrucksstark:
(Bsp: kontravariante Typargumente in Java, C#)
Programmierer muß Sprachstandard lesen und verstehen und dazu Konzepte (z.B. aus Vorlesung) kennen

Fachmännisches Programmieren

- Hardware: wer Flugzeug/Brücke/Staudamm/... baut, kann (und darf) das auch nicht allein nach etwas Selbststudium und mit Werkzeug aus dem Baumarkt
- Software: der (Bastel-)Prototyp wird oft zum Produkt, der Bastler zum selbsternannten Programmierer,
- bei einigen Programmiersprachen ähnlich
 - BASIC (1964) (Kemeny, Kurtz) to enable students in fields other than science and math. to use computers

- PHP (1994) (Rasmus Lerdorf) Personal Home Page Tools (like Perl, more limited, simple) <https://www.php.net/manual/en/history.php>
- Python (van Rossum) 1999 *Computer Programming for Everybody* proposal <https://www.python.org/doc/essays/cp4e/>

Legacy-Sprachen: ECMA-Script (Javascript)

- \approx LISP (1960) (Funktionen als Daten, keine stat. Typ.)
- ursprüngliches Ziel: Software soll auf Endgerät laufen
- technisches Problem: Gerätebenutzer versteht/beherrscht seinen Computer/Betriebssystem nicht (z.B. will oder darf keine JRE installieren)
- stattdessen zwingt man die Werbe-Zielpersonen auf Browser mit Javascript-Engine (der Browser ist das OS)
- das steckt z.B. Google viel Geld hinein: <https://v8.dev/docs/turbofan> (der JIT-Compiler rät die fehlenden Typen)
- zusätzliche Motivation: billige Front-End-Programmierer auch für Arbeiten am Back-End (Server)

Aktuelle Entwicklungen: JS

- ECMA-Script 6 übernimmt viele Konzepte moderner (funktionaler) Programmierung, u.a.
 - let (block scope), const (single assignment)
 - destructuring (pattern matching)
 - tail calls (ohne Stack)

<https://tc39.es/ecma262/>

- ... was ist mit Microsoft? Die haben auch viel Geld und clevere Leute? — Ja: <http://www.typescriptlang.org/>
TypeScript adds *optional types*, classes, and modules to JavaScript.

Personen: Luke Hoban, Anders Hejlsberg, Erik Meijer, ...

Legacy-Sprachen: PHP

- Facebook ist (ursprünglich) in PHP implementiert
- deswegen steckt Facebook viel Geld in Technologien (a.k.a. Work-Arounds) für diese Sprache
aus ebenfalls verständlichen Gründen :
 - für Kunden (d.h. Werbekunden): Antwortzeiten der Webserver
 - für Betreiber: Entwicklungs- und Betriebskosten

Aktuelle Entwicklungen: PHP

- HHVM: Hip Hop Virtual Machine
<https://github.com/facebook/hhvm/blob/master/hphp/doc/bytecode.specification>
- Hack <http://hacklang.org/> Type Annotations, Generics, Nullable types, Collections, Lambdas, ...

Julien Verlaquet: *Facebook: Analyzing PHP statically*, 2013, <http://cufp.org/2013/julien-verlaquet-facebook-analyzing-php-statically.html>
vgl. Neil Savage: *Gradual Evolution*, Communications of the ACM, Vol. 57 No. 10, Pages 16-18, <http://cacm.acm.org/magazines/2014/10/178775-gradual-evolution/fulltext>

Aktueller: Web Assembly

- a new portable, size- and load-time-efficient format suitable for compilation to the web.
<http://webassembly.org/>
- d.h., Programme in vernünftigen (d.h. typsicheren) Sprachen schreiben (statt JS), nach WASM kompilieren und im Browser ausführen
- das gabe es alles schon? Natürlich:
 - Java → Bytecode (class files) (1996),
 - Pascal → P(ortable)-Code (1973)
- formale Spezifikation (typsichere Kellermaschine) <https://webassembly.github.io/spec/core/index.html>

Die Zukunft: Typen für Ressourcen

<https://www.rust-lang.org/>

... a systems programming language that ... prevents segfaults and guarantees thread safety.

- jedes Datum hat genau einen *Eigentümer*, man kann Daten *übernehmen* und *ausborgen*,
- *statisch* garantiert: für jedes Datum $x : T$ gibt es
 - exactly one mutable reference ($\&\text{mut } T$),
 - one or more references ($\&T$)

<https://github.com/rust-lang/rust-wiki-backup/blob/master/Note-research.md#type-system>,

lineare Logic (Girard 1987), siehe <https://www.cs.cmu.edu/~fp/courses/linear/lectures/lecture16.html>

Die Zukunft: Datenabhängige Typen

<https://idris-lang.org/> ... aspects of a program's behaviour can be specified *precisely* in the type.

- elementare Bausteine:
 - Daten: 42, "foo", $(x, y) \Rightarrow x+y$, Typen: bool, int
- Kombinationen (Funktionen):
 - Datum \rightarrow Datum, Bsp. $(x, y) \Rightarrow x+y$
 - Typ \rightarrow Typ, Bsp. `List<T>`
 - Typ \rightarrow Datum, Bsp. `Collections.<String>sort () [10pt]`
 - Datum \rightarrow Type, (*data-*)*dependent type*, Bsp. Vektoren
`data Vec : Nat -> Type -> Type`

```
(++) : Vec p a -> Vec q a -> Vec (p+q) a  
head : Vect (S p) a -> a    -- S = Nachfolger
```

Nicht reguläre Typen

- ein rekursiver polymorpher Typ heißt *regulär*, wenn die Typ-Argumente bei Rekursion gleich bleiben

```
data List a = Nil | Cons a (List a),
```


... sonst *nicht regulär*
- ```
data List a b = Nil | Cons a (List b a)
```
- ```
data Tree a = Leaf a | Branch (Tree (a,a))
```
- Anwendung: Implementierung von `containers:Data.Sequence` <https://hackage.haskell.org/package/containers-0.6.2.1/docs/Data-Sequence.html#t:Seq>

15 Zusammenfassung

Themen Methoden zur Beschreibung der

- Syntax: reguläre Ausdrücke, kontextfreie Grammatiken
- Semantik: operational, denotational, axiomatisch
- Konzepte:
 - Typen,
 - Namen (Deklarationen), Blöcke (Sichtbarkeitsbereiche)
 - Ausdrücke und Anweisungen (Wert und Wirkung),
 - Unterprogramme (als Daten)
 - Polymorphie (statisch, dynamisch)
- Wechselwirkungen der Konzepte
- Paradigmen: imperativ, funktional, objektorientiert

Sprachen kommen und gehen, Konzepte bleiben.

Wie weiter? (LV)

Anwendung und Vertiefung von Themen PPS z.B. in VL

- Programmverifikation
u.a. axiomatische Semantik imperativer Programme
- Compilerbau
 - Realisierung der Semantik durch
 - * Interpretation, * Transformation (Kompilation)
 - abstrakte und konkrete Syntax (Parser)
- Constraint-Programmierung
- Computermusik (Programme für Klänge und Musik)
- Symbolisches Rechnen (Transformation von Syntax-Bäumen, Semantik: z.B. Polynome, Funktionen, geometrische Konstruktionen)

Übungen

1. Colin McMillen, Jason Reed, and Elly Fong-Jones, 2011: *Programming Language Checklist* https://famicol.in/language_checklist.html You appear to be advocating a new ... programming language. Your language will not work. Here is why:...

Diskutieren Sie die dort genannten Eigenschaften (Pragmatik). Wie sind sie (in dieser VL) definiert? (Semantik)

2. zu Folie *nicht reguläre Typen*:
 - (a) Objekte vom Typ `List Bool Int` konstruieren,
 - (b) vom Typ `Tree Bool`
 - (c) `Data.Sequence`: wie sieht eine Folge der Länge 10 intern aus? (der *finger tree* unter dem `Seq`-Konstruktor)
3. Typescript, Rust, Idris, Agda, ... ausprobieren