Programmieren für Mathematiker WS2017/18

Dozent: Prof. Dr. Wolfgang Walter

13. August 2018

Inhaltsverzeichnis

Kapitel I

allgemeine Informationen

Eine Programmiersprache ist lexikalisch, syntaktisch und semantisch eindeutig definiert. Eine Compiler übersetzt die Programmiersprache in Maschinensprache . Ein Interpreter arbeitet das Programm dann ab. Ein Laufzeitsystem stellt grundlegende Operationen und Funktionen zur Verfügung.

Bereiche der Informatik

Die Informatik untergliedert sich in 4 Bereiche:

- Technische Informatik
- Praktische Informatik
- Theoretische Informatik
- Angewandte Informatik

Die Technische Informatik beschäftigt sich mit der Konstruktion der Hardware, zum Beispiel der Datenleitungen, um Informationen durch das Internet zu transportieren. Wichtige Firmen sind hier: Intel, Globalfoundries und Infineon.

Die Praktische Informatik beschäftigt sich mit der Software, also Betriebssystem, Compiler, Interpreter und so weiter. In alltäglicher Software findet sich rund 1 Fehler in 100 Zeilen Quelltext. In wichtiger Software, also Raketen, Betriebssysteme, ..., ist es nur 1 Fehler pro 10.000 Zeilen Code.

Die Theoretische Informatik beschäftigt sich mit Logik, formalen Sprachen, der Automatentheorie, Komplexität von Algorithmen, ...

Die Angewandte Informatik beschäftigt sich mit der Praxis, dem Nutzer, der Interaktion zwischen Mensch und Maschine, ...

Maßeinheiten und Größenordnungen

Ein bit ist ein Kunstwort aus "binary" und "digit". Es kann nur 2 Werte speichern: 0 und 1

Ein nibble ist eine Hexadezimalziffer, bündelt also 4 bits und kann damit 16 Werte annehmen: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E und F.

Ein byte bündelt 2 nibble, also 8 bit. Er ist die gebräuchlichste, direkt addressierbare, kleinste Speichereinheit. Weitere Speichergrößen sind:

Name	**Anzahl byte**	**Name**	**Anzahl byte**
1 KB	10^3	1 KiB	$2^{10} = 1.024$
1 MB	10^6	1 MiB	$2^{20} = 1.048.576$
1 GB	10^9	1 GiB	$2^{30} = 1.073.741.824$
1 TB	10^{12}	1 TiB	2^{40}
1 PB	10^{15}	1 PiB	2^{50}
1 EB	10^{18}	1 EiB	2^{60}

Der ROM ("read-only-memory") speichert wichtige Informationen auch ohne Strom, wie zum Beispiel die Uhrzeit, Informationen über die Festplatte, ... Er ist nicht mehr änderbar, außer durch Belichtung.

Der RAM ("random-access-memory") ermöglicht den Zugriff auf alle Adressen, insbesondere im Hauptspeicher.

Kapitel II

Zahldarstellungen

Basis-Konvertierung ganzer Zahlen

Die Notation $[9]_{10}$ bedeutet, dass man die Zahl 9 im Zehner-System betrachtet. Es gilt also $[9]_{10} = [1001]_2$ und $[10]_{10} = [1010]_2$.

Um eine Zahl von einer gegebenen Basis in eine Zielbasis b zu konvertieren, so teilt man immer wieder durch b und notiert den Rest als nächste Ziffer von hinten nach vorne. Am Beispiel von $[57]_{10}$ ins Zweier-System sieht das so aus:

$$\begin{aligned}
\frac{57}{2} &= 28 \text{ Rest } 1 \Rightarrow \text{ letzte Ziffer der Binärdarstellung}\\
\frac{28}{2} &= 14 \text{ Rest } 0 \Rightarrow \text{ vorletzte Ziffer der Binärdarstellung}\\
\frac{14}{2} &= 7 \text{ Rest } 0\\
\frac{7}{2} &= 3 \text{ Rest } 1\\
\frac{3}{2} &= 1 \text{ Rest } 1\\
\frac{1}{2} &= 0 \text{ Rest } 1
\end{aligned}$$

Also gilt: $[57]_{10} = [111001]_2$.

Die umgekehrte Richtung verläuft ähnlich:

$$\begin{array}{r}
111001 \quad : \quad 1010 \quad = \quad 101 \;\; \text{R} \;\; 111\\
-1010\\
\hline
01000\\
-00000\\
\hline
10001\\
-01010\\
\hline
111
\end{array}$$

Also $[111001]_2$ durch $[10]_{10} = [1010]_2$ gleich $[101 \text{ Rest } 111]_2 = [5 \text{ Rest } 7]_{10} \Rightarrow [57]_{10}$.

Von Basis 2 in Basis 4, 8 oder 16 ist dann ganz einfach: $[111100101]_2$

- Zweiergruppen von hinten nach vorne zusammenzählen: $[13211]_4$
- Dreiergruppen von hinten nach vorne zusammenzählen: $[745]_8$
- Vierergruppen von hinten nach vorne zusammenzählen: $[1E5]_{16}$

Basis-Konversion gebrochener Zahlen

Festkommadarstellung (nur Betrag der Zahl, ohne Vorzeichen):

Gewichte	B^k	B^{k-1}	...	B^1	B^0	.	B^{-1}	B^{-2}	...	B^{-l}
Ziffern	m_k	m_{k+1}	...	m_{-1}	m_0	.	m_1	m_2	...	m_l

Also: $\sum\limits_{i=k}^{l} m_i \cdot B^{-i}$.

Die Konvertierung des ganzzahligen Anteils vor dem "." läuft wie gehabt. Um den gebrochenen Anteil zu konvertieren, multipliziert man wiederholt mit der Zielbasis b und nimmt den jeweiligen ganzzahligen Anteil als Nachkommaziffern (von links nach rechts). Mit dem gebrochenen Anteil macht man weiter. Wir wollen die Zahl $[0.625]_{10}$ ins Zweiersystem konvertieren:

$$\begin{aligned} 0.625 \cdot 2 &= \mathbf{1}.25 \\ 0.25 \cdot 2 &= \mathbf{0}.5 \\ 0.5 \cdot 2 &= \mathbf{1} \end{aligned}$$

Also gilt: $[0.625]_{10} = [0.101]_2$.

Wieder anders herum:

$$\begin{aligned} 0.101 \cdot 1010 &= \mathbf{110}.010 \\ 0.010 \cdot 1010 &= \mathbf{10}.100 \\ 0.100 \cdot 1010 &= \mathbf{101}.0 \end{aligned}$$

Also gilt $[0.101]_2 = [0.110|10|101]_2 = [0.625]_{10}$.

Jetzt wollen wir $[0.1]_{10}$ ins Zweiersystem konvertieren:

$$\begin{aligned} 0.1 \cdot 2 &= \mathbf{0}.2 \\ 0.2 \cdot 2 &= \mathbf{0}.4 && (1) \\ 0.4 \cdot 2 &= \mathbf{0}.8 \\ 0.8 \cdot 2 &= \mathbf{1}.6 \\ 0.6 \cdot 2 &= \mathbf{1}.2 \\ 0.2 \cdot 2 &= \mathbf{0}.4 && (2) \end{aligned}$$

Wie man sieht, sind die Zeilen (1) und (2) gleich, das heißt, diese Konvertierung wird unendlich lange laufen. Also: $[0.1]_{10} = [0.0\overline{0011}]_2$. Aber es muss gelten: $[0.1]_{10} \cdot [10]_{10} = [1]_{10}$. Aber es stimmt: $[0.0\overline{0011}]_2 \cdot [1010]_2 = [0.\overline{1}]_2 = [1]_2$.

Entsprechend gilt:

$$\begin{aligned}
[0.2]_{10} &= [0.\overline{0011}]_2 \\
[0.3]_{10} &= [0.01\overline{0011}]_2 \\
[0.4]_{10} &= [0.011\overline{0011}]_2 \\
[0.5]_{10} &= [0.1]_2 \\
[0.6]_{10} &= [0.1\overline{0011}]_2 \\
[0.7]_{10} &= [0.1011\overline{0011}]_2 \\
[0.8]_{10} &= [0.11\overline{0011}]_2 \\
[0.9]_{10} &= [0.111\overline{0011}]_2
\end{aligned}$$

Problem: Rundungen schon bei $\frac{1}{10} \Rightarrow$ falsche Nachkommastellen. Die Lösung sind hier Gleitkommazahlen.

Gleitkommazahlen

Gleitkommazahlen werden auch Fließkommazahlen, Gleitpunktzahlen, Fließpunktzahlen oder floating-point-numbers genannt.

Das Gleitkommaformat $R = (b, l, \underline{e}, \overline{e})$ besteht aus

- einer Basis b
- einer Mantissenlänge l
- einem Exponentenbereich von $\underline{e}$ bis $\overline{e}$.

Eine Gleitkommazahl ist entweder 0 oder $x = (-1)^s \cdot m \cdot b^e$ mit

- Vorzeichenbit $s \in \{0, 1\}$
- Mantisse $m = [0.m_1 m_2 m_3 ... m_l]_b$ mit Mantissenziffern $m_i \in \{0, 1, 2, ..., b-1\}$
- $e \in \{\underline{e}, \underline{e}+1, \underline{e}+2, ..., \overline{e}\}$

Schauen wir uns das Beispiel $R(2, 3, -1, +2)$ an. Eine solche Zahl benötigt 1 bit für s, 2 bits für e und 3 bits für m.

$m =$ **0.**	**111**	**110**	**101**	**100**	**011**	**010**	**001**	**000**
$e = -1$	$\frac{7}{16}$	$\frac{6}{16}$	$\frac{5}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	0
$e = 0$	$\frac{14}{16}$	$\frac{12}{16}$	$\frac{10}{16}$	$\frac{8}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{2}{16}$	0
$e = 1$	$\frac{28}{16}$	$\frac{24}{16}$	$\frac{20}{16}$	$\frac{16}{16}$	$\frac{12}{16}$	$\frac{8}{16}$	$\frac{4}{16}$	0
$e = 2$	$\frac{56}{16}$	$\frac{48}{16}$	$\frac{40}{16}$	$\frac{32}{16}$	$\frac{24}{16}$	$\frac{16}{16}$	$\frac{8}{16}$	0

Es gibt also auch mehrere Darstellungen für eine Zahl! Die Cyan eingefärbten Zahlen können auch anders dargestellt werden.

Grüne Zahlen sind sogenannte normalisierte Gleitkommazahlen , ihre erste Mantissenziffer ist $\neq 0$. Die roten Zahlen sind denormalisierte Gleitkommazahlen : Ihre ersten Mantissenziffer ist $m_1 = 0$ und ihr Exponent $e = \underline{e}$. Da das erste Mantissenbit häufig eine 1 ist, wird angenommen, dass das erste Mantissenbit eine 1 ist und wird deswegen nicht gespeichert (hidden bit). Das sorgt dafür, dass bei 3 bit Genauigkeit mit 4 bit Genauigkeit gerechnet werden kann. Ist das erste Mantissenbit eine 0, gibt es dafür eine spezielle Exponentenkennung.

Ein Zahlenstrahl mit diesen Zahlen ist besonders dicht um 0, aber ab 2 werden die Abstände sehr groß.

Die größte darstellbare Zahl ist $x_{max} = 0.1111...1 = (1 - b^l) \cdot b^{\overline{e}}$.
Der kleinste darstellbare normalisierte Betrag ist $x_{min,N} = 0.10000...0 = b^{\underline{e}-1}$.
Der kleinste darstellbare denormalisierte Betrag ist $x_{min,D} = 0.0000...1 = b^{\underline{e}-l}$.

Doch es gibt eine Probleme:

- absolute/relative Fehler bei Zahlen, die zwischen 2 darstellbaren Zahlen liegen $\Rightarrow$ Rundungen bei nahezu jeder Rechnung!
- Grundrechenarten können nicht darstellbare Zahlen erzeugen

Rundung

Eine Rundung ist eine Funktion $O : \mathbb{R} \to$ Gleitkomma-Raster R.

Eine Rundung O hat folgende Eigenschaften:

1. $O(x) = x$ wenn $x \in R$
2. $x, y \in \mathbb{R}$ mit $x < y \Rightarrow O(x) < O(y)$
3. $O(-x) = -O(x)$, nur manche Rundungen haben diese Eigenschaft

Es gibt verschiedene Rundungsmodi:

- "to nearest": zur nächsten Gleitkommazahl, wenn 2 Gleitkommazahlen gleich weit weg sind, wird abwechselnd auf- und abgerundet
- "trancation": Abschneiden der Nachkommastellen $\Rightarrow$ betragskleiner runden
- "augmentation": zusätzliche Stellen hinzufügen $\Rightarrow$ betragsgrößer runden
- "upward": nach oben runden
- "downward": nac unten runden

Wenn O eine Rundung mit einem Rundungsmodus, also $O \in \{\text{Rundungsmodi}\}$, ist und $\circ$ eine Grundrechenart, also $\circ \in \{+, -, \cdot, \div\}$, dann gilt für eine Gleitkommaoperation $\odot$

$$x, y \in R : x \odot y := O(x \circ y)$$

Auslöschung in Summen von Gleitkommazahlen tritt auf, wenn die Größenordnung der exakten Summe wesentlich kleiner ist als die Größenordnung der Summanden (bzw. der Zwischenergebnisse).

Kapitel III

Grundstrukturen von Algorithmen

Begriffe

Eine Sequenz sind einzelne Anweisungen hintereinander.
Eine Selektion ist eine Verzweigung.
Eine Repetition ist eine Wiederholung.

Variablen und Daten

Wir sehen uns einen Biertrinker an, der nach dem Genuss noch ein paar Besorgungen machen muss. Dabei ergeben sich folgende Variablen

- Variable **Durst** von Typ *LOGICAL*
- Variable **Geld** von Typ *INTEGER*
- Variable **PreisDerBesorgung** von Typ *INTEGER*
- Variable **Rest** von Typ *INTEGER*
- Variable **Bierpreis** von Typ *INTEGER*
- Variable **WirtschaftAnnehmbar** von Typ *LOGICAL*
- Variable **Autofahrer** von Typ *LOGICAL*
- Variable **AlkoholGrenzwert** von Typ *REAL*
- Variable **AlkoholVergiftungsWert** von Typ *REAL*

Schleifen

In Fortran gibt es 4 Arten von Schleifen :

- Endlosschleife
- Schleife mit Anfangsbedingung
- Schleife mit Endbedingung
- Zählschleife

Bei einer Endlosschleife wird der Anweisungsblock innerhalb der Schleife unendlich lange ausgeführt:

```
do
 Anweisung1
 Anweisung2
 Anweisung3
```

```
end do
```

Bei einer Schleife mit Anfangsbedingung wird der Anweisungsblock nur ausgeführt, wenn die Anfangsbedingung wahr ist. Ist sie wahr, so wird der Block ausgeführt und anschließend überprüft, ob die Anfangsbedingung wieder wahr ist. Ist die Anfangsbedingung nicht wahr, so wird die Schleife nicht ausgeführt.

```
do while(Anfangsbedingung)
 Anweisung1
 Anweisung2
 Anweisung3
end do
```

Hat die Schleife eine Endbedingung , so wird der Anweisungsblock auf jedem Fall 1-mal ausgeführt. Erst dann wird überprüft, ob die Endbedingung wahr ist. Ist sie das, wird die Schleife **verlassen**. Ist sie falsch, so wird die Schleife erneut ausgeführt.

```
do
 Anweisung1
 Anweisung2
 Anweisung3
 if(Endbedingung) exit
end do
```

Eine Zählschleife in Fortran ist ähnlich wie in anderen Programmiersprachen konzipiert, aber die Syntax ist deutlich verschieden. Eine Zählschleife besitzt eine Zählvariable (die auch in der Schleife benutzt werden kann, aber nicht geändert werden sollte), die von einer Anfangszahl mit bestimmter Schrittweite solange hochgezählt (oder bei negativer Schrittweite heruntergezählt) wird, bis die Zählvariable die Endzahl erreicht.

```
do i = anfang, ende, schrittweite
 Anweisung1
 Anweisung2
 Anweisung3
end do
```

Bitte beachten:

- Der Zustand der Zählvariable `i` vor der Schleife geht verloren, auch wenn die Schleife 0-mal läuft.
- Die Zählvariable `i` darf im Inneren der Schleife nicht geändert werden.
- Der Endzustand der Zählvariable `i` ist nach der Schleife nicht definiert.
- Ausdrücke werden zu Beginn genau 1-mal (vor der ersten Iteration) berechnet und sind dann fest.
- Die Anzahl der Iterationen ist: $N = \max\left\{0, \mathtt{nint}\left(\frac{\mathtt{ende} - \mathtt{anfang} + \mathtt{i}}{\mathtt{i}}\right)\right\}$

Einfache Syntax

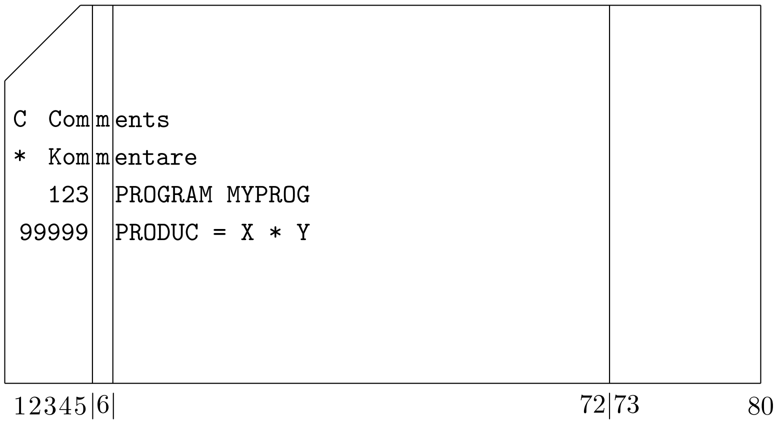
Es gibt einen zulässigen Fortran-Zeichensatz. Dieser umfasst zum Beispiel die Buchstaben A-Z und a-z, sowie 0-9, Sonderzeichen und Operatoren.

Nun einige lexikalische Einheiten (Symbole/Tokens):

- Keywords sind nicht reserviert, Variablen können also auch nach Keywords benannt werden.
- Identifiers (Namen) haben eine Länge von maximal 63 Zeichen. Variablen können Buchstaben, Zahlen und auch den Unterstrich enthalten.
- Literale (Konstanten): 3 ⇒ INTEGER, 2.876 ⇒ REAL, .TRUE. ⇒ LOGICAL, "Hallo" ⇒ STRING
- Labels (Marken): 00000 ... 99999 sind Sprungmarken, die mit `GOTO 99999` erreicht werden können.
- Separatoren (Trennsymbole) sind: `()`, `/`, `/(/)`, `[]`, `=`, `=>`, `:`, `::`, `,`, `;` und `%`.
- Operatoren sind: `+`, `-`, `*`,`/`, `**`, `//`, `==`, `<=`, `<`, `/=`, `>`, `>=`, `.NOT.`, `.OR.`, `.AND.`, `.EQV.` und `.NEQV.`.

In den alten Quellformen bis vor Fortran-90, also insbesondere Fortran-66 und Fortran-77 waren

- Namen maximal 6 Zeichen lang und
- Lochkarten 80 Zeichen breit.



Ab Fortran-90 wurden neue Quellformen entwickelt, das heißt:

- Zeilen sind nun maximal 132 Zeichen lang
- Kommentare beginnen mit "!" und gehen bis zum Zeilenende
- Eine neue Zeile ist eine neue Anweisung, außer die letzte Zeile endet mit "&"
- "&" am Zeilenende bedeutet, dass die nächste nicht-Kommentar und nicht-Leerezeile die Anweisung fortsetzt.
- Die Fortsetzung darf mit "&" beginnen.
- Es sind maximal 39 Fortsaetzungszeilen möglich
- Leerzeichen sind signifkant ⇒ alle lexikalischen Tokens sind am Stück zu schreiben.
- Groß- und Kleinschreibung ist nicht signifikant in Namen und Keywords

Kapitel IV

Ein- und Ausgabe

Anhang