

Einführung in die Programmiertechnik

Darstellung von Informationen

Informatik

- Begriff entstand in den 60er Jahren aus dem Begriffen Information und Automatik
 - English: üblicherweise „computer science“ („informatics“: science of information), „information technology“
- Definition der GI (Gesellschaft für Informatik), :
Informatik ist die Wissenschaft von der systematischen und automatisierten Verarbeitung von Information. Sie erforscht grundlegende Verfahrensweisen der Informationsverarbeitung und allgemeine Methoden ihrer Anwendung in den verschiedenen Bereichen. Für diese Aufgaben wendet die Informatik vorwiegend formale und ingenieurmäßig orientierte Techniken an.

Informatik: Unterteilung nach Gumm

- **Technische Informatik:** Konstruktion von Rechnern, Speicherchips, ...
 - Hardware
- **Praktische Informatik:** Programmierung
 - Software-Entwicklung, Software-Qualitätssicherung, Betriebssysteme, Datenbanken, Compilerbau, ...
- **Theoretische Informatik:** mathematische Grundlagen
 - formale Sprachen, Automatentheorie, Validierung und Verifikation
- **Angewandte Informatik:** Einsatz von Rechnern im täglichen Leben
 - Büroautomatisierung, medizinische Informatik, Bioinformatik, Wirtschaftsinformatik, ...

Informatik: Verschiedene Sichten nach Balzert

- Position 1: **Mathematisch-logische Orientierung**
 - Computer arbeiten nach formalen Prinzipien, Programmierer müssen sich also die Formalismen aneignen und reale Probleme in den Formalismus übertragen.
- Position 2: **Ingenieurwissenschaftliche Orientierung**
 - Informatiker sollten wie Ingenieure arbeiten, also systematische Vorgehensweisen (Planung, Modellierung) auf konkrete praktische Probleme anwenden.
- Position 3: **Evolutionäre Orientierung**
 - Eine exakte Formalisierung des praktischen Problems ist oft nicht möglich. Deshalb sind Versuche formalen, geplanten Vorgehens gescheitert. Man muss das akzeptieren und in seiner Arbeit vorhersehen.
- Position 4: **Partizipative Orientierung**
 - Computer erfüllen bestimmte Funktionen innerhalb menschlicher Tätigkeitsbereiche. Der Informatiker muss sich also mit dem sozialen Kontext des Arbeitsplatzes, der Arbeitsverteilung usw. beschäftigen.

Information (nach Broy)

- Trennung zwischen äußerer Darstellung einer Nachricht und dem Inhalt der Nachricht
 - „Es regnet“
 - «Дождь идёт»
 - 下雨
- *Information* nennen wir den abstrakten Gehalt („Bedeutungsinhalt“, „Semantik“) eines Dokuments, einer Aussage, Beschreibung, Anweisung, Nachricht oder Mitteilung. Die äußere Form der Darstellung nennen wir die *Repräsentation* (konkrete Form der Nachricht).

Darstellung von Information

- Viele Formen der Repräsentation
 - verabredete Zeichen („Signale“)
 - gesprochenes Wort („akustische Darstellung“)
 - taktile Reize, Wärmewahrnehmung
 - Zeichnungen
 - Zeichenfolgen (geschriebenes „Wort“, „Text“)
- Den (häufig nur gedanklichen) Übergang von der Repräsentation zur abstrakten Information, die Deutung der Repräsentation, nennen wir *Interpretation*.
- Prozess der Interpretation im Gehirn: Verstehen
 - formal schwierig erfassbar
 - deshalb: mathematisch-logische Sicht begnügt sich mit der Abbildung der Repräsentation auf mathematisches Modell; Verständnis des Modells muss unabhängig erworben werden

Bits

- kleinstmögliche Einheit der Information
- Antwort auf Frage, die zwei Möglichkeiten zulässt
 - ja oder nein
 - wahr oder falsch
 - schwarz oder weiß
 - hell oder dunkel
 - links oder rechts
- **Kodierung** der Information: Festlegung einer Repräsentation (eines Codes)
 - Darstellung von Bits: Binärcode
 - Zwei Zeichen: oft **0** und **1** (auch: 0 und L)
 - Repräsentation im Computer: oft durch Ladungen
 - 0 = ungeladen, 1 = geladen
 - 0 = 0 Volt, 1 = 5 Volt
 - 0 = unmagnetisiert, 1 = magnetisiert

Bitfolgen

- Repräsentation von mehr als zwei Antworten: mehrere Bits
 - Aus welcher Himmelsrichtung (Ost, Süd, Nord, West) weht der Wind?
 1. Weht der Wind aus einer der Richtungen Nord oder Ost (ja/nein)?
 2. Weht der Wind aus einer der Richtungen West oder Ost (ja/nein)?
 - 0 0: Süd, 0 1: West, 1 0: Nord, 1 1: Ost
- Mit jedem Bit verdoppelt sich die Zahl der Antworten
 - 3 Bits: Windrichtung mit Zwischenrichtungen
 - 0 0 0 = Süd, 0 0 1 = Südwest, 0 1 0 = West, 0 1 1 = Nordwest,
 - 1 0 0 = Nord, 1 0 1 = Nordost, 1 1 0 = Ost, 1 1 1 = Südost
- Es gibt genau 2^N mögliche Bitfolgen der Länge N

Hexziffern

- Lange Bitfolgen für den Menschen schwer erfassbar
- Konvention: Bitfolgen werden zu Vierergruppen zusammengefasst

0000 = 0	0100 = 4	1000 = 8	1100 = C
0001 = 1	0101 = 5	1001 = 9	1101 = D
0010 = 2	0110 = 6	1010 = A	1110 = E
0011 = 3	0111 = 7	1011 = B	1111 = F

- Statt 0100 1111 0110 0001 kann man dann 4F61 schreiben

Zahlendarstellung in einem Positionssystem

- Verallgemeinerung des Dezimalsystems
- Zahlensystem zur Basis N: N Ziffern mit Werten von 0..N-1
 - Ziffer: Symbol aus einer Symbolmenge
- Seien $a_m a_{m-1} a_{m-2} \dots a_2 a_1 a_0$ Ziffernfolge
- Durch Ziffernfolge dargestellte Zahl:
$$a_m N^m + a_{m-1} N^{m-1} + a_{m-2} N^{m-2} + \dots + a_2 N^2 + a_1 N + a_0$$

Verbreitete Positionssysteme

- Dezimalsystem: Basis 10, Ziffern 0,1,2,3,4,5,6,7,8,9
- Binärsystem: Basis 2, Ziffern 0,1
- Oktalsystem: Basis 8, Ziffern 0,1,2,3,4,5,6,7
- Hexadezimalsystem: Basis 16, Ziffern 0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
 - oft statt Großbuchstaben auch Kleinbuchstaben verwendet
- Bei Uneindeutigkeit wird Basis als Index zugefügt
 - $13_{16} = 19_{10} = 23_8 = 10011_2$

Bytes und Worte

- Rechner verarbeiten Bits immer in Gruppen:
 - übliche Gruppengröße: 8 Bits, 16 Bits, 32 Bits, 64 Bits
 - heutzutage üblicherweise Vielfache von 8
- **Byte**: Zusammenfassung von 8 Bit
 - exakte Begriffsdefinition variiert
 - C-Standard: „Byte“ == „Datenmenge, die in Wert von Typ char gespeichert wird“
 - Zur Klarstellung wird auch der Begriff „Oktett“ (octet) verwendet
- **Worte**: „Natürliche“ Gruppierung von Bits im Rechner
 - Historisch (16-Bit-Rechner): Wort = 2 Byte, Doppelwort = 4 Byte, Quadwort = 8 Byte
 - Begriffsdefinition stark unheitlich; auf einem 32-Bit-Rechner bezeichnet man oft mit „Wort“ 4 Byte (Halbwort: 2 Byte)

Dateien

- Folgen von Bytes auf einem Speichermedium (Festplatte, CD-ROM, ...)
 - „Folge“ nur im einfachsten (verbreitetsten) Fall, i.allg. auch flexiblere Gruppierung anstelle sequentieller Folgen möglich
- Organisation von mehreren Dateien in einem Computer üblicherweise durch hierarchische Dateinamen
- Interpretation des Inhalts einer Datei durch Programm oder Nutzer
 - Erkennung der Art der Daten durch Namenskonventionen (etwa: Dateinamensanhänge) oder Dateiformat (Festlegung spezifischer Bytes etwa am Dateianfang)

Dateigrößen

- Üblich: Zählung der Bytes „in“ einer Datei
 - i.d.R. ohne Zählung der Bytes in Dateinamen, sonstigen Dateiattributen
- Abkürzung B für Bytes: 245**B**
- Große Zahlen (z.B.: 460146819B) schlecht handhabbar
 - Lösungsidee: metrisches System (kB = 1000B, MB = 1000000B)
- Aber: Speicher oft in Zweierpotenzen organisiert
 - Sektorgröße auf Festplatten: 512B
 - Größe eines modernen Speichermoduls: 1073741824B

Dateigrößen

- Lösung: $2^{10} = 1024 \approx 1000$
- Einheitenpräfixe ähnlich dem metrischen System (Addendum zu IEC 60027-2)
 - $2^{10} = 1024 = \text{ki}$ (= kibi)
 - $2^{20} = 1024 \times 1024 = \text{Mi}$ (= Mebi)
 - $2^{30} = 1024 \times 1024 \times 1024 = \text{Gi}$ (= Gibi)
 - $2^{40} = 1024 \times 1024 \times 1024 \times 1024 = \text{Ti}$ (= Tebi)
 - $2^{50} = 1024 \times 1024 \times 1024 \times 1024 \times 1024 = \text{Pi}$ (= Pebi)
 - $2^{60} = 1024 \times 1024 \times 1024 \times 1024 \times 1024 \times 1024 = \text{Ei}$ (= Exbi)
- Verwendung: kiB (Kibibyte), MiB (Mebibyte), GiB (Gibibyte)
- Historisch auch noch die mathematisch falsche Verwendung üblich
 - kB als wird Kilobyte ausgesprochen und als 1024 Byte verstanden
 - u.U. auch metrisch richtig verwendet (etwa bei Festplattengrößen: 1GB dann 10^9B)

Ausblick

- Darstellung spezifischer Daten
 - Zahlen
 - Text
 - Programme
 - Bilder und Musik