

# Rechnersysteme

Marcel Waldvogel

## Übersicht

- Wer bin ich?
- Die Vorlesung
- Heutige Vorlesung: Daten
- Bücher

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 2

## Wer bin ich?

- 33 Jahre, verheiratet, 2-jähriger Sohn
- ETH Zürich
- Uptime Object Factory
- Washington University in St. Louis
- IBM Zurich Research Laboratory
- Interessen:
  - System-/Netzwerktuning
  - Netze für Demokratie, Freiheit, Sicherheit

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 3

## Die Vorlesung

- Ohne Leistungsnachweis
- Grundlagen für die Zukunft
  - Rechner: Grundlagen, Komponenten
  - Netzwerke: Übersicht
  - Systemsoftware: Grundlagen
- Erwartungen?

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 4

## Inhalt

- Daten
- Schaltkreise/-netze
- von-Neumann-Prinzip
- RISC/CISC
- Assembler (2 Vorlesungen)
- Speicherhierarchie
- Betriebssysteme (2 Vorlesungen)
- Systemsoftware
- Rechnernetze (2 Vorlesungen)
- Verteilte Systeme

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 5

## Daten im Rechner

- Was ist Information?
- Was sind Daten?
- Wie werden Daten gespeichert?
- Wie werden Daten übermittelt?
- Datenkompression

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 6

## Darstellung von Daten

- **Codierung: Grundvoraussetzung für elektronische Speicherung, Verarbeitung und Übermittlung von Information**
  - Geeignete Darstellung von Information, Nachrichten, Daten
- **Quellencodierung**
  - Kompression
  - Verschlüsselung
- **Kanalcodierung**
  - Redundanz

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 7

## Speichern von Information

- **Unterste Ebene: Binäre Logik**
  - 0 und 1
  - wahr und falsch
  - ein und aus
- **Darstellung**
  - Spannung
  - Ladung
  - Magnetisierung
  - Loch
  - Ton
  - Helligkeit
- **Analog vs. digital**

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 8

## Binäre Logik

- **Variablen**
- **Unäre und binäre Operatoren**
  - Wahrheitstablen
- **Arithmetik**
  - Vereinfachungen
- **Codewörter unterschiedlicher Länge**
  - Heute typisch: 8, 16, 32, 64

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 9

## Zahlen

- **Sequenz von Bits**
  - Dezimal vs. oktal/hexadezimal vs. BCD
- **Vorzeichen**
  - Einer-, Zweierkomplement
- **Arithmetik**
  - Überlauf
  - Übertrag
- **Fließkommazahlen**
  - Vorzeichen, Mantisse, Exponent

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 10

## Zeichen und Text

- **Buchstaben**
  - Gross und klein
- **Zahlen**
- **Satzzeichen**
  - , . ? !
- **Umlaute, nationale Sonderzeichen**
  - ü é ç
- **Sonderzeichen**
  - @\$%^&\*()\_-=+ \ | { } [ ] " ' / ~ < >
- **Nicht-lateinische Zeichen**
  - Indien, China, Japan, Korea; Hieroglyphen
- **Steuer-, Kontrollzeichen**

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 11

## I Speak English

- **Baudot (Telex)**
  - 5 Bit; "Shift"
- **ASCII**
  - 7 Bit+Parität
  - Länderspezifische Variationen
- **EBCDIC**
- **ISO Latin-1 (ISO-8859-1)**
  - 8 Bit, aufbauend auf ASCII
- **Unicode**
  - 16 bzw. 32 Bit
  - UTF-8
- **HTML**
  - Namen für Symbole

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 12

## Information

- **Informationsmenge**
  - Entropie
  - Beispiel: Wochentage
- **Information vs. Daten**
- **Ist das Information?**
  - "Ich bin ein Huhn."
  - "Der nächste Satz ist wahr. Der obige Satz ist falsch."

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 13

## Kompression

- **Reduktion des Umfangs der Daten unter Beibehaltung der Information**
- **Ziel: Speicherplatz und Übertragungszeit sparen**
- **Quellencodierung**
  - Nachrichten werden bei der Quelle codiert, bevor sie über einen Übertragungskanal geschickt werden
  - Wissen über die Art der Quelle hilfreich

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 14

## Kompression (Forts.)

- **Häufigere Nachrichten erhalten kürzere Codesequenzen als weniger häufigere**
  - Nie auftretende Nachrichten erhalten keine Codesequenz
- **Vereinfachung oder Kompression?**
  - 001b vs. "Montag"
  - UTF-8
- **Verlust**
  - Buchhaltung vs. Multimedia

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 15

## Informationstheorie

- **Wahrscheinlichkeitsrechnung, Statistik**

- **Quelle  $S$  mit Symbolen  $s_1, \dots, s_n$**

- $s_i$  hat **Wahrscheinlichkeit**  $P(s_i)=p_i$

- $p_i \geq 0; \sum p_i = 1$

- **Informationsgehalt von  $s_i$ :**  $I(s_i) = \log_2 \frac{1}{p_i}$

- **Gesamtentropie der Quelle:**

$$H(S) = \sum p_i I(s_i) = \sum p_k \log \frac{1}{p_k}$$

- **Voraussetzung: Quelle stationär und gedächtnislos**

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 16

## Quellencodierungstheorem

- **Shannon 1948:**

- Eine Quelle mit Entropie  $H_Q$  kann mit beliebig hoher Wahrscheinlichkeit mit einem Code mit Entropie  $H_K$  minimal grösser als  $H_Q$  codiert werden, aber nicht mit einem Code mit kleinerer Entropie als  $H_Q$

- **Kompression besser als Entropie unmöglich**

- **Tricks?**

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 17

## Laufängencodierung

- **Engl.: Run-Length Coding**

- **Fügt vor jedem Symbol einen Wiederholungswert ein (Zahlen binär codiert)**

- AAAABBBCCCC -> 4A3B5C (komprimiert)
- ABCBAC -> 1A1B1C1B1A1C (expandiert)

- **Verfeinerung: Länge von unkomprimierbaren Folgen (z.B. Länge+100)**

- ABCAAAABBB -> 103ABC4A3B

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 18

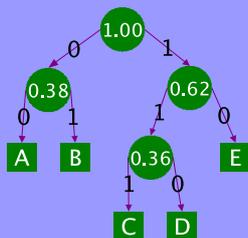
# Huffman-Kodierung

- **Codierende Bitfolge abhängig von der Wahrscheinlichkeit eines Symbols**
  - **Statisch**
  - **Z.B. für Faxübermittlung**
- **Rekursiver Aufbau eines Codebaumes**
  - **Seltenste zwei Symbole werden zusammengefasst und durch Anhängen von 0 bzw. 1 unterschieden**
  - **Eindeutiges Ende der Sequenz (Symbole nur in den Blättern des Baumes)**

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 19

## Huffman: Beispiel

Symbole	A	B	C	D	E
Wahrscheinlichkeit	0.16	0.22	0.14	0.12	0.36



Symbole	A	B	C	D	E
Codesequenz	00	01	111	110	10

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 20

## Lempel-Ziv

- **Lempel und Ziv 1977:**  
Beliebige schon gesehene Teilfolgen werden zur Übertragung weiterer Symbole genutzt: Die längste mögliche Teilfolge, die bereits gesehen wurde, wird gesucht und ihre Anfangsposition und Länge übertragen
  - **Dynamische Kompression**
  - **"Deflate" (LZ77+Huffman), z.B. (g)zip, zlib**
- **Lempel und Ziv 1978:**  
Nachricht wird in Teilfolgen aufgeteilt, die in einem Wörterbuch abgelegt werden. Das jeweils längste passende Wort im Wörterbuch wird gesucht und seine Nummer übertragen

Marcel Waldvogel, IBM Zurich Research Laboratory, Universität Konstanz, 15.10.2001, 21

## Weitere Kompressionsverfahren

- **LZW [Welch 1984]**
  - **Automatischer Baufbau für LZ78**
- **Burrows-Wheeler Transform [1994]**
  - **Schaut (indirekt) auch in die Zukunft**
    - **Generiert rotierte Versionen der Nachricht**
    - **Sortiert diese Versionen**
    - **Komprimiert die letzte Spalte der sortierten Matrix**
  - **Z.B. in bzip2**