



Ergänzungs-Studienangebot der Mediendidaktik für
Lehramtsstudenten
Dr. rer. nat. Harald Sack
Institut für Informatik
FSU Jena

Sommersemester 2004

Informatik der digitalen Medien

1 21.04.2004 – Vorlesung Nr. 2 3 4 5 6 7 8 9 10 11 12
13
14

2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Informatik der digitalen Medien

2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Warum Kodierung?
- Multimediale Daten im Computer
 - Zeichen, Text, Bild, Audio und Video
- Zahlen
 - Binärkodierung von Dezimalzahlen
 - Negative Zahlen im 2er-Komplement
 - Binärkodierung von Gleitkommazahlen

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Informatik der digitalen Medien

2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Zeichencodes
 - Samuel Morse's Telegraficode
 - Fernschreiber und der Baudot-Code
 - ASCII-Code und nationale Erweiterungen
 - Unicode – ein Code für alle Sprachen und Schriften

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Informatik der digitalen Medien

2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Warum Kodierung?
- Multimediale Daten im Computer
- Zahlen
- Zeichencodes

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Warum Kodierung?
 - „natürliche Information“ ist analog
 - z.B. **Musik**
 - an bestimmter Stelle liegen Töne mit bestimmter
 - Tonhöhe
 - Lautstärke
 - Charakteristik
 - etc...

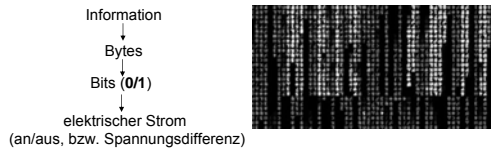


Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Warum Kodierung?

- „natürliche Information“ ist analog
- aber
- Darstellung im Computer ist **digital**



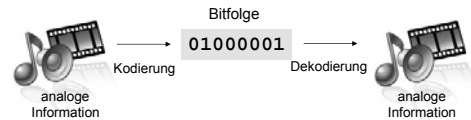
Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Warum Kodierung?

○ Was ist Kodierung?

- Vorgang der **Umwandlung** einer Information
 - aus einer bestimmten Darstellung (Ausprägung)
 - in eine **andere Darstellung**
- ohne (bzw. mit zu tolerierendem) Informationsverlust
- Rückwandlung wird als Dekodierung bezeichnet



Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Informatik der digitalen Medien

2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Warum Kodierung?
- **Multimediale Daten im Computer**
- Zahlen
- Zeichencodes

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Multimedia?

- Information, die mit Hilfe **mehrerer, verschiedenartiger Medien** zur Darstellung kommt
- z.B.
 - Text und Bild
 - Text und Bild und Audio/Video
 - etc.



Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Medientypen

- **Text**
 - Kodierung **alphanumerischer Information**
 - Buchstaben/Ziffern
 - Kodierung z.B. mit
 - verschiedenen Alphabeten, z.B. αβγδεϕηθ...
 - Morse-Code, z.B. ...-.-.-.-.-
 - ASCII-Code
 - Unicode



Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Medientypen

- **Grafik**
 - Kodierung von **Bildinformation**
 - Raster-/Vektorgrafik
 - Klassifikation nach Komplexität der verwendeten Farbpalette
 - Monochrom bis
 - Echtfarben
 - Abhängig von menschlicher Wahrnehmung



Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Medientypen

○ Audio

- Kodierung von **akustischer Information**
- Sprache / Musik / Geräusche
- Abhängig von menschlicher Wahrnehmung

- Wichtiger Unterschied zu Text/Grafik:

⇒ Wiedergabe von Audio-Daten **muss in Echtzeit erfolgen!**



Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07144 Jena, E-Mail: sack@minet.uni-jena.de

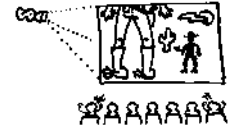
Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Medientypen

○ Video / Animation

- Kodierung von Bild- und Tonfolgen
- Wiedergabe muss ebenfalls in **Echtzeit** erfolgen
- Abhängig von menschlicher Wahrnehmung



Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07144 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Medientypen

○ Zeitunabhängige Medien

- Zeitkomponente bei Aufzeichnung/Wiedergabe ohne Bedeutung
- z.B. Text und Grafik
- „diskrete“ Medien

○ Zeitabhängige Medien

- Information verändert sich mit der Zeit
- Gehalt der Einzelinformation nicht signifikant
- Gesamtinformation erschließt sich erst aus zeitlichem Ablauf
- Zeitkomponente kritisch bei Aufzeichnung/Wiedergabe
- z.B. Audio und Video

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07144 Jena, E-Mail: sack@minet.uni-jena.de

Informatik der digitalen Medien

2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Warum Kodierung?
- Multimediale Daten im Computer
- **Zahlen**
- Zeichencodes

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07144 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Zahlen

○ Dezimalsystem

- Basis **10**
- Ziffern 0,1,2,3,4,5,6,7,8,9
- 1428 = 1·1.000 + 4·100 + 2·10 + 8·1
= 1·10³ + 4·10² + 2·10¹ + 8·10⁰

○ Binär-/Dualsystem

- Basis **2**
- Ziffern 0,1
- 1001 = 1·2³ + 0·2² + 0·2¹ + 1·2⁰
= 1·8 + 0·4 + 0·2 + 1·1
= **9 (dezimal)**

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07144 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Zahlen

- Verallgemeinerung: **p**-adische Zahlen
- Basis **p**, p ∈ ℕ
- Ziffern **0,1,...,p-1**
- Darstellung einer Zahl k ∈ ℕ als p-adische Zahl

$$k = \sum_{i=0}^n c_i \cdot p^i, \quad c_i \in \{0, 1, \dots, p-1\}$$

- Bsp: p=10, Ziffern: 0,1,2,3,4,5,6,7,8,9

$$1234 = 1 \cdot 10^3 + 2 \cdot 10^2 + 3 \cdot 10^1 + 4 \cdot 10^0$$

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07144 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Zahlen

○ Hexadezimalsystem

- in der Informatik weit verbreitet
- Basis: $p=16$
- Ziffern: **1,2,3,4,5,6,7,8,9,A,B,C,D,E,F**
- Bsp.:

$$\begin{aligned} 2F9 &= 2 \cdot 16^2 + 15 \cdot 16^1 + 9 \cdot 16^0 \\ &= 2 \cdot 256 + 15 \cdot 16 + 9 \cdot 1 \\ &= 512 + 240 + 9 \\ &= 761 \end{aligned}$$

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Zahlen

○ Umrechnung Dualsystem - Dezimalsystem

- $101001 = 1 \cdot 2^5 + 1 \cdot 2^3 + 1 \cdot 2^0$
 $= 1 \cdot 32 + 1 \cdot 8 + 1 \cdot 1$
 $= 41$ (dezimal)
- $53 = 1 \cdot 32 + 1 \cdot 16 + 0 \cdot 8 + 1 \cdot 4 + 0 \cdot 2 + 1 \cdot 1$
 $= 1 \cdot 2^5 + 1 \cdot 2^4 + \quad + 1 \cdot 2^2 + \quad + 1 \cdot 2^0$
 $= 110101$

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Zahlen

○ Umrechnung Dualsystem - Dezimalsystem

- $53 = 1 \cdot 32 + 1 \cdot 16 + 0 \cdot 8 + 1 \cdot 4 + 0 \cdot 2 + 1 \cdot 1$
 $= 1 \cdot 2^5 + 1 \cdot 2^4 + \quad + 1 \cdot 2^2 + \quad + 1 \cdot 2^0$
 $= 110101$

○ Einfacher Algorithmus:

$53:2 =$	26	Rest	1	↑ Leserichtung
$26:2 =$	13	Rest	0	
$13:2 =$	6	Rest	1	
$6:2 =$	3	Rest	0	
$3:2 =$	1	Rest	1	
$1:2 =$	0	Rest	1	

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Zahlen

○ Darstellung im Computer

- Bytes (1 Byte = 8 Bit)
- Bit (Wert: 0 oder 1)
- Bsp.:
- 53 als **8-Bit Dualzahl** = 00110101

Stelle	7	6	5	4	3	2	1	0
Bitwert	0	0	1	1	0	1	0	1
Dezimalwert	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Zahlen

○ Einfache Arithmetik

○ Addition

$$\begin{array}{r} 19 \quad 00010011 \\ +9 \quad 00001001 \\ \hline \quad \quad \quad 11 \\ \hline 28 \quad 00011100 \end{array}$$

○ Subtraktion

$$\begin{array}{r} 19 \quad 00010011 \\ -9 \quad 00001001 \\ \hline \quad \quad \quad 1 \\ \hline 10 \quad 00001010 \end{array}$$

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Multimediale Daten im Computer

○ Negative Dualzahlen

○ 1-Komplement:

- für negative Zahlen werden alle Bits **invertiert** (umgedreht) (1 wird 0 und 0 wird 1)
- und ein zusätzliches Bit (**Vorzeichenbit**) =1 gesetzt

○ Bsp.:

- 53 als **8-Bit Dualzahl** = 00110101
- -53 als **8-Bit Dualzahl im 1-Komplement** = 11001010

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

Multimediale Daten im Computer

Negative Zahlen

- Bsp.:
 - 53 als **8-Bit Dualzahl** = 00110101
 - 53 als **8-Bit Dualzahl im 1-Komplement** = 11001010

V	6	5	4	3	2	1	0
1	1	0	0	1	0	1	0
	2^6	2^5	2^4	2^3	2^2	2^1	2^0

Stelle (oben rechts), Bitwert (Mitte rechts), Dezimalwert (unten rechts), Vorzeichenbit (unten links)

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de


Grundlagen der Digitalisierung – Datenrepräsentation im Computer

Multimediale Daten im Computer

Negative Zahlen

- Problem bei 1-Komplement:
 - Es gibt 2 verschiedene Kodierungen für 0:
 - 00000000 = 11111111
 - Addition und Subtraktion mit negativen Zahlen ergibt falsches Ergebnis

z.B. $17 + (-12)$

$$\begin{array}{r} 17 \quad 00010001 \\ -12 \quad 11110011 \\ \hline = \quad 00000100 = 4 \end{array}$$


Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

Multimediale Daten im Computer

Negative Zahlen

- 2-Komplement:**
 - für negative Zahlen werden alle Bits **invertiert** (umgedreht) (1 wird 0 und 0 wird 1)
 - zur invertierten Zahl wird **1 hinzuaddiert**
 - ein zusätzliches Bit (**Vorzeichenbit**) = 1 wird gesetzt
- Bsp.:
 - 53 als **8-Bit Dualzahl** = 00110101
 - 53 als **8-Bit Dualzahl im 1-Komplement** = 11001010
 - 53 als **8-Bit Dualzahl im 2-Komplement** = **11001011**

Bemerkung: anstelle (2) kann auch vor der Invertierung von der zu invertierenden Zahl 1 subtrahiert werden.

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de


Grundlagen der Digitalisierung – Datenrepräsentation im Computer

Multimediale Daten im Computer

Negative Zahlen

- 2-Komplement:
 - es gibt nur noch eine Null: 00000000
 - Addition und Subtraktion funktionieren korrekt, auch für negative Zahlen im 2-Komplement

z.B. $17 + (-12)$

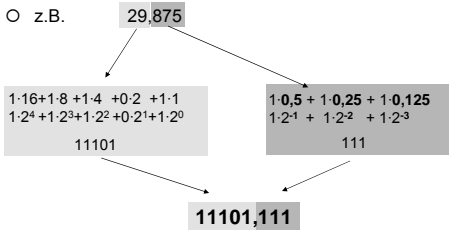
$$\begin{array}{r} 17 \quad 00010001 \\ -12 \quad 11110100 \\ \hline = \quad 00000101 = 5 \end{array}$$


Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

Multimediale Daten im Computer

Gleitkommazahlen



Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

Multimediale Daten im Computer

Gleitkommazahlen

- das Komma steht immer an einer anderen Stelle.....
- kann man das Komma fixieren?

$$\begin{array}{l} 29,875 = 2,9875 \cdot 10^1 \\ 3,14159 = 3,14159 \cdot 10^0 \\ 12342,7 = 1,23427 \cdot 10^4 \end{array}$$

- Fixiere Komma an erster Stelle (**Normalisierung**)
 $1,23427 \cdot 10^4$

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- **Multimediale Daten im Computer**
 - **Gleitkommazahlen**
 - Fixiere Komma an erster Stelle (**Normalisierung**)

$$1,23427 \cdot 10^4$$

- Trennung in **Mantisse** und **Exponent**

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- **Multimediale Daten im Computer**

- **Binäre Gleitkommazahlen**

$$29,875 = 1101,111$$

Normalisierung

$$1,101111$$

Mantisse

$$1,101111$$

$$2^3$$

Exponent

$$11$$

Vorzeichenbit

V	M	M	M	M	M	M	M	M	M	E	E	E	E	
0	1	1	0	1	1	1	1	0	0	0	0	0	1	1

16-Bit Gleitkommazahl:

Vorzeichenbit:

1 Bit

Mantisse:

10 Bit

Exponent:

5 Bit

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- **Multimediale Daten im Computer**
 - **Binäre Gleitkommazahlen**
 - üblich ist heute 32/64-Bit Gleitkommadarstellung
 - IEEE 754 Floating Point Representation
 - 1 Bit Vorzeichen
 - 8/11 Bit Exponent
 - 23/52 Bit Mantisse
 - Beispielanwendung

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- **Multimediale Daten im Computer**
 - **Zahlendarstellung - Zusammenfassung**
 - Zahlen werden im Computer stets als **Binärzahl** repräsentiert
 - Unterscheide:
 - Einfache, positive Binärdarstellung
 - Negative Binärdarstellung
 - 1-Komplement
 - 2-Komplement
 - normalisierte Gleitkommadarstellung

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Informatik der digitalen Medien

2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

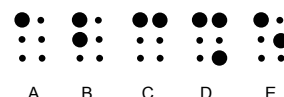
- Warum Kodierung?
- Multimediale Daten im Computer
- Zahlen
- **Zeichencodes**

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- **Textkodierung**

- Alphabet: a,b,c,d,e,...,A,B,C,D,E,...,0,1,2,...
- Z.B.: Brailleschrift



Louis Braille (1809-1852)

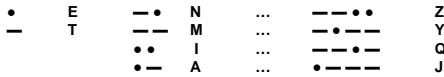
- Zeichen werden in 3x2-BinärMatrix kodiert

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ Morsecode



Samuel Morse
1791 to 1872

Samuel Morse (1791-1872)

- binärer Zeichencode (1840) (Strom an/aus)
- Buchstaben (1 – 4 Signale)
- Zahlen (5 Signale)
- Interpunktion (6 Signale)
- Zeichen/Wortgrenzen werden durch Pausen markiert
- Zeichenlänge abhängig von mittlerer Häufigkeit des Zeichens

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ Baudot-Code

○ Problem:

- Morsecode wegen unterschiedlicher Zeichenlänge zu kompliziert für **Fernschreiber**

- daher konstante Zeichenlänge: **5 Bit** (Emile Baudot's Fernschreibmaschine, 1874)

- **Problem:** reicht aber nur für 32 verschiedene Zeichen



○ Idee: Doppelbelegungen

- spezielles **Steuerzeichen** schaltet zwischen Ziffern und Buchstaben um



Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ ASCII-Code

- **(American Standard Code for Information Interchange)**
 - in 50er Jahre gab es keinen Kodierungsstandard für **Computer**
 - **1961** (Robert Bemer, IBM) entwickelt **7-Bit** Kodierung
 - basiert auf 7-Bit FIELDATA Code
 - 99 Zeichen (Ziffern, Großbuchstaben, Steuersymbole)
 - ECMA belegt Rest mit Kleinbuchstaben
 - 1963 erstmals von ANSI standardisiert
 - 1974 ISOI-646 Standard
 - von IBM aber erst 1981 eingesetzt

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ ASCII-Code

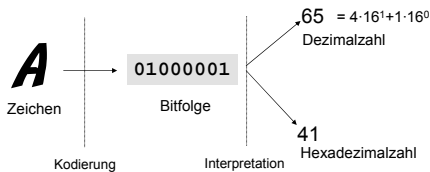
ASCII value	Character	Control character	ASCII value	Character	ASCII value	Character	ASCII value	Character
000	(null)		010	(space)	014	n	026	v
001	(start of heading)		011	SOH	015	O	027	w
002	(start of text)		012	STX	016	P	028	x
003	(end of text)		013	ETX	017	Q	029	y
004	(end of transmission)		014	EOT	018	R	030	z
005	(quit)		015	ENOQ	019	S	031	[
006	(cancel)		016	ACN	020	T	032	\
007	(bell)		017	BEL	021	G	033]
008	(backspace)		018	BS	022	H	034	^
009	(tab)		019	HT	023	I	035	_
010	(line feed)		020	LF	024	J	036	`
011	(form feed)		021	FF	025	K	037	a
012	(page feed)		022	PF	026	L	038	b
013	(carriage return)		023	CR	027	M	039	c
014	(shift out)		024	SO	028	N	040	d
015	(shift in)		025	SI	029	O	041	e
016	(data link escape)		026	DLE	030	P	042	f
017	(escape)		027	DC1	031	Q	043	g
018	(file separator)		028	DC2	032	R	044	h
019	(group separator)		029	DC3	033	S	045	i
020	(record separator)		030	DC4	034	T	046	j
021	(unit separator)		031	NAK	035	U	047	k
022	(separator)		032	SYN	036	V	048	l
023	(end of transmission)		033	ETB	037	W	049	m
024	(cancel)		034	CAF	038	X	050	n
025	(end of transmission)		035	EM	039	Y	051	o
026	(end of transmission)		036	ESC	040	Z	052	p
027	(end of transmission)		037	FS	041	[053	q
028	(end of transmission)		038	GS	042	\	054	r
029	(end of transmission)		039	RS	043]	055	s
030	(end of transmission)		040	SS	044	^	056	t
031	(end of transmission)		041	OS	045	_	057	u
032	(end of transmission)		042	BS	046	`	058	v
033	(end of transmission)		043	OS	047	a	059	w
034	(end of transmission)		044	OS	048	b	060	x
035	(end of transmission)		045	OS	049	c	061	y
036	(end of transmission)		046	OS	050	d	062	z
037	(end of transmission)		047	OS	051	e	063	[
038	(end of transmission)		048	OS	052	f	064	\
039	(end of transmission)		049	OS	053	g	065]
040	(end of transmission)		050	OS	054	h	066	^
041	(end of transmission)		051	OS	055	i	067	_
042	(end of transmission)		052	OS	056	j	068	`
043	(end of transmission)		053	OS	057	k	069	a
044	(end of transmission)		054	OS	058	l	070	b
045	(end of transmission)		055	OS	059	m	071	c
046	(end of transmission)		056	OS	060	n	072	d
047	(end of transmission)		057	OS	061	o	073	e
048	(end of transmission)		058	OS	062	p	074	f
049	(end of transmission)		059	OS	063	q	075	g
050	(end of transmission)		060	OS	064	r	076	h
051	(end of transmission)		061	OS	065	s	077	i
052	(end of transmission)		062	OS	066	t	078	j
053	(end of transmission)		063	OS	067	u	079	k
054	(end of transmission)		064	OS	068	v	080	l
055	(end of transmission)		065	OS	069	w	081	m
056	(end of transmission)		066	OS	070	x	082	n
057	(end of transmission)		067	OS	071	y	083	o
058	(end of transmission)		068	OS	072	z	084	[
059	(end of transmission)		069	OS	073	{	085	\
060	(end of transmission)		070	OS	074		086	}
061	(end of transmission)		071	OS	075	~	087	

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ ASCII-Code



Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ ASCII-Code

○ Problem:

- 7 Bit ausreichend für 128 Zeichen
- International existieren aber viele Umlaute und Sonderzeichen

○ Lösung:

- **ISO 8859-x** Standard,
 - 8-Bit ASCII-Kodierung mit nationalen Erweiterungen (Umlaute)
 - 0-127 identisch mit Standard-ASCII
 - 128-159 seltene Steuerzeichen
 - **160-255 nationale Erweiterungen**

Informatik der digitalen Medien
Dr. rer. nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ ASCII-Code

○ nationalen Erweiterungen

- ISO-8859-1 Westeuropa, Amerika, Australien, Afrika (ISO-8859-15)
- ISO-8859-2 Osteuropa (ISO-8859-16)
- ISO-8859-3 Esperanto und Maltesisch
- ISO-8859-4 Baltisch, Grönland, Lappland
- ISO-8859-5 Bulgarien, Mazedonien, Russisch, Serbien, Ukraine
- ISO-8859-6 Arabisch (ohne Persisch/Urdu)
- ISO-8859-7 Griechenland
- ISO-8859-8 Hebräisch
- ISO-8859-9 Island, Türkei
- ISO-8859-10 Grönland, Lappland
- ISO-8859-11 Thai
- ISO-8859-12 Indien
- ISO-8859-13 Baltikum
- ISO-8859-14 Gälisch, Walisisch

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ ASCII-Code

○ Problem:

- 8 Bit ausreichend für 256 Zeichen
- Chinesische, japanische, koreanische oder indische Schriftzeichen lassen sich damit nur schwer repräsentieren
- Bsp.: chinesische Schriftzeichen in Japan
 - Gakashu Kanji: 1006 Zeichen (Grundschule)
 - Juyo Kanji 1945 Zeichen (offizielle Dokumente, Zeitung)
 - Jinmei-yo Kanji 285 Zeichen (Namen)

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ ASCII-Code

○ Problem:

- 8 Bit ausreichend für 256 Zeichen
- Chinesische, japanische, koreanische oder indische Schriftzeichen lassen sich damit nur schwer repräsentieren
- Unterschiedliche Laufrichtungen
 - Bsp.: deutsch **hallo**
 - hebräisch מלך טוב
- Multilinguale Dokumente

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ ASCII-Code

○ Problem:

- 8 Bit ausreichend für 256 Zeichen
- Chinesische, japanische, koreanische oder indische Schriftzeichen lassen sich damit nur schwer repräsentieren
- Multilinguale Dokumente

○ Lösung:

- mehr Bits pro Zeichen, um größere Zeichenvorräte kodieren zu können

 **16-Bit Unicode**

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ Unicode

- ab 1984, ISO 10646 Standard seit 1992
- ursprünglich 16-Bit, dann 21 (32)-Bit Kodierung
- ermöglicht **multilinguale Textverarbeitung**
- genutzt werden 17 Ebenen (planes) mit je 65.536 Zeichen
- potenziell aber 2.147.483.648 Zeichen möglich
- **Basic Multilingual Plane (BMP, Unicode 3.2.0, 2002)**
 - 49194 ausgewiesene Zeichen
 - unterstützt fast alle gebräuchlichen modernen Schreibsysteme

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ Unicode

○ **Basic Multilingual Plane (BMP)**

- kann in 16-Bit UTF-16 kodiert werden
- Schreibweise: **U+xxxx₁₆**
- BMP umfasst
 - 10236 Buchstaben
 - 27786 **CJK**-Unihan-Zeichen (vereinheitlichte chinesische, japanische und koreanische Schrift)
 - 11172 koreanische Hangul-Zeichen
 - 8515 Kontrollsymbole
- **Supplementary Multilingual Plane (SMP)**
 - archaische und nicht mehr genutzte Schriftsysteme, Musik, Mathe

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-07444 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ Unicode

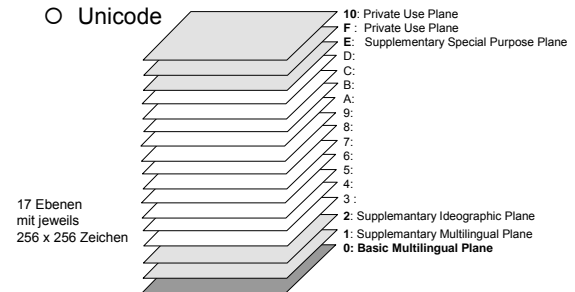
- Supplementary Ideographic Plane (SIP)
 - CJK-Erweiterungen
 - Seltene und unübliche chinesische Schriftzeichen
- Supplementary Special Purpose (SSP)
 - Kontrollsymbole
 - Protokollanweisungen
- Private Use Planes (PUP)

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ Unicode



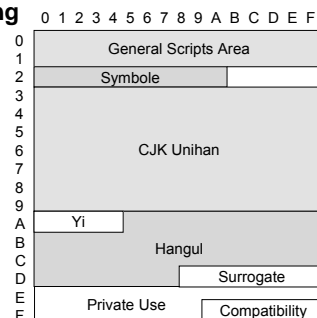
Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ Unicode

● BMP



Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ Unicode

● BMP – General Script Area

- 0-127 entspricht ASCII-Kodierung (Kompatibilität) (U+0000 – U+007F)
- 0-255 entspricht ISO-8859-1 (U+0000 – U+00FF)
- umfasst Lateinschrift, Griechisch, Kyrrilisch, Hebräisch, Arabisch, Devanagari, Bengali, Äthiopisch, Cherokee, Kmehr, Runen, Thai, Laotisch, Malayalam, Mongolisch, Tibetisch, Telegu, Georgisch, Ogham, Gurmukhi, Oriya, Tamil, Sinhala, Thaana, etc.

vgl. www.unicode.org

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

○ Textkodierung

○ Unicode – Universal Transformation Formats

- UTF-8
 - 8-Bit Unicodevariante, identisch mit ASCII
- UTF-16
 - 16-Bit Kodierung der BMP-Unicode-Zeichen
- UTF-32
 - 32-Bit Unicode Kodierung nach ISO 10646
 - UCS-4 (Universal Multiple Octet Coded Character Set)

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Informatik der digitalen Medien

2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Warum Kodierung?
- Multimediale Daten im Computer
- Zahlen
- Zeichencodes

Informatik der digitalen Medien
Dr rer nat Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- **Multimediale Daten im Computer**
 - **Zeichenkodierung - Zusammenfassung**
 - Die Anzahl der jeweils darstellbaren Zeichen hängt von der Länge der verwendeten Zeichenkodierung ab
 - z.B. Baudot: 5 Bit, ASCII: 7 Bit
 - Entwicklung der Zeichenkodierung
 - Morse Code
 - Baudot Code
 - ASCII
 - Unicode

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07144 Jena, E-Mail: sack@informatik.uni-jena.de

Informatik der digitalen Medien

2. Grundlagen der Digitalisierung –
Datenrepräsentation im Computer
 - **Literatur**
 - Ch. Meinel, H. Sack:
WWW– Kommunikation, Internetworking, Web-Technologien,
Springer, 2004.
 - P.A. Henning:
Taschenbuch Multimedia,
3. Aufl., Fachbuchverlag Leipzig, 2003.
 - R. Gilliam:
**Unicode Demystified –
A Practical Programmer's Guide to the Encoding Standard**
Addison Wesley, 2003.
 - G. Brookshear:
„1.5 The Binary System“,
in Computer Science – an Overview,
Addison Wesley, NY, 2003, pp. 43–56.
 - The Unicode Consortium:
<http://www.unicode.org>

Informatik der digitalen Medien
Dr rer nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2.4, D-07144 Jena, E-Mail: sack@informatik.uni-jena.de