

Algorithmen und Datenstrukturen

Algorithmen und Datenstrukturen

Foliensatz, 03.11.05

Wintersemester 2005/2006

Information

Information

- Modell
 - Menge von Attributen
 - sowie Menge der Beziehungen zwischen Attributen.
- Attribute und Beziehungen repräsentieren
 - Merkmale
 - andere wesentlichen Eigenschaften eines realen Systems
 - Vereinfachungen gegenüber System nötig, um Modell handhabbar zu halten

Information

- Modellanalyse
 - ermittelt Werte (Merkmalsausprägungen) der Attribute
 - Zahlen
 - Farben
 - Mengen von Personen
 - Namen
 - Formen
 - Definition
 - *Zustand (state)* eines Modells
 - Belegung der Attribute des Modells mit konkreten Werten
 - Gesamtheit der Ausprägungen aller Attribute eines Modells

Information

- Beispiele für Zustände
 - Statische Größen
 - Dynamische Größen
 - (gr. dynamikós=mächtig, kräftig, stark; lat. dynamice, frz. (18. Jh.)
dynamique="Lehre von der Bewegung bzw. Kraft")
 - Planetenmodell
 - Zustand jedes Körpers wird beschrieben durch Masse, seine Geschwindigkeit (Betrag und Richtung der Bewegung) sowie seine Ortskoordinaten zu einer bestimmten Zeit
 - Masse des Körpers bleibt konstant
 - Geschwindigkeit ändert sich dynamische
 - Position ändert sich dynamische
 - Zeit selbst spielt Sonderrolle
 - Parameter von Ort und Geschwindigkeit

Information

- Modellcharakter
 - Newtonsche Mechanik
 - konstante, unveränderte Masse eines bewegten Körpers
 - spezielle Relativitätstheorie
 - genaueres Modell
 - Masse abhängig von Geschwindigkeit
 - nur in sehr geringem Umfang
 - Newtons Modell 'nur' Annäherung an reale Verhältnisse
 - nach heutiger Kenntnis sehr gute Annäherung

Information

- Beispiele
 - Planetenmodell
 - Planet erreicht Ausgangsposition (oder kommt ihr am nächsten)
 - „Jahr“ für jeweiligen Planeten
 - verstrichene Zeit
 - Position der anderen Planeten
 - Zwei-Körper-Modell
 - Keplerschen Gesetze
 - Fahrstrahl überstreicht in gleichen Zeiten gleiche Fläche
 - Quadrate der Umlaufzeiten verhalten sich wie dritten Potenzen der großen Halbachsen
 - Baustatik
 - Träger einer Brücke wird solange verstärkt, bis Tragfähigkeit der Brücke erreicht
 - Pythagoräischer Lehrsatz
 - invariante Beziehung: $a^2 + b^2 = c^2$

Information

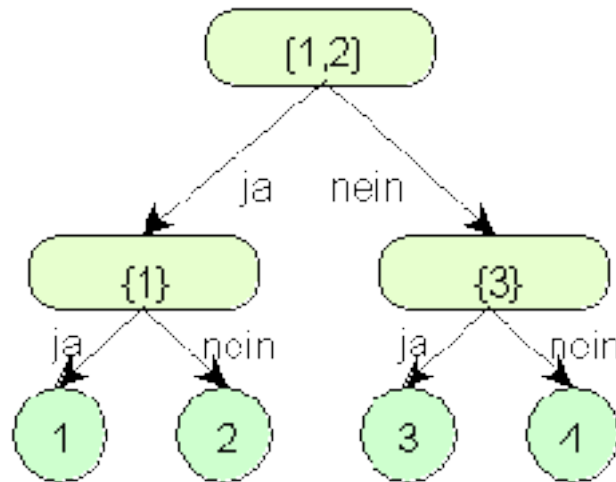
- **Information** (*information*)
 - Wissen über Zustand eines Systems
 - umgangssprachlich sehr viel unschärfer
 - Wissen nicht zählbar, also auch Information nicht
 - Information kann mathematisch behandelt werden
 - Informationsmenge I
 - Entropie H
 - Nicht-Informationsmenge (Unwissen über Zustand eines Systems)
 - $I = H$

Informationstheorie

- **Informationstheorie**
- Information mathematisch untersuchen/modellieren
 - *Shannon: A mathematical Theory of Communication 1959*
 - Übertragung von Information über Kanal
 - Empfänger stellt Fragen
 - Sender darf nur mit „Ja“ oder „Nein“ antworten
 - Quantifizierung der Information
 - optimale Anzahl von Fragen, um den Zustand des Systems zu ermitteln
 - Beispiel
 - Modell mit vier Zuständen
 - Ermittle Zustand aus Menge $\{1, 2, 3, 4\}$
 - frage, ob aktueller Zustand in Menge $\{1,2\}$
 - „Ja“: frage, ob aktueller Zustand 1
 - „Nein“: frage, ob aktueller Zustand 3
 - Gewichtetes Mittel: $4 \cdot 2 / 4 = 2$ Fragen.

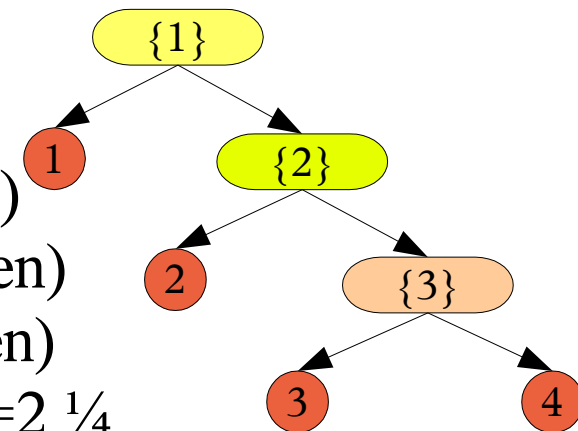
Informationstheorie

- Fragefolge anhand 'Codebaum' aufzeichnen



- Beispiel: Alternatives Frageschema

- In Zustand 1? (Für einen Zustand eine Frage)
- In Zustand 2? (Für einen Zustand zwei Fragen)
- In Zustand 3? (Für zwei Zustände drei Fragen)
- Gewichtete Summe: $1 \cdot 1/4 + 1 \cdot 2/4 + 2 \cdot 3/4 = 9/4 = 2 \frac{1}{4}$



Informationstheorie

- Informationsgehalt
 - minimale mittlere Anzahl von Fragen, die nötig ist, um den Zustand des Systems zu ermitteln
 - Mathematische Definition des Begriffs Informationsmenge Entropie H
 - Die **ideelle Informationsmenge** eines Systems mit N Zuständen, in welchem der Zustand i mit der Häufigkeit p_i auftritt, ist:

$$H = \sum_i -p_i \cdot \log_2 p_i$$

- Dabei ist \log_2 der Logarithmus zur Basis 2
- relative Häufigkeit p_i für alle Zustände gleich $p = p_i$: $H = -\log_2 p$
- 4 Zustände, $p_i = 0,25 = 2^{-2}$: $H = -\log_2 2^{-2} = 2$

Informationstheorie

- Informationsgehalt

- Beispiel

- System mit 8 Zuständen, relative Häufigkeiten $p_1 = p_2 = 1/4$; $p_3 = p_4 = 1/8$; $p_5 = p_6 = p_7 = p_8 = 1/16$.
- für Zustände 1 und 2 jeweils zwei Fragen
- für Zustände 3 und 4 jeweils drei Fragen
- sonst jeweils vier Fragen

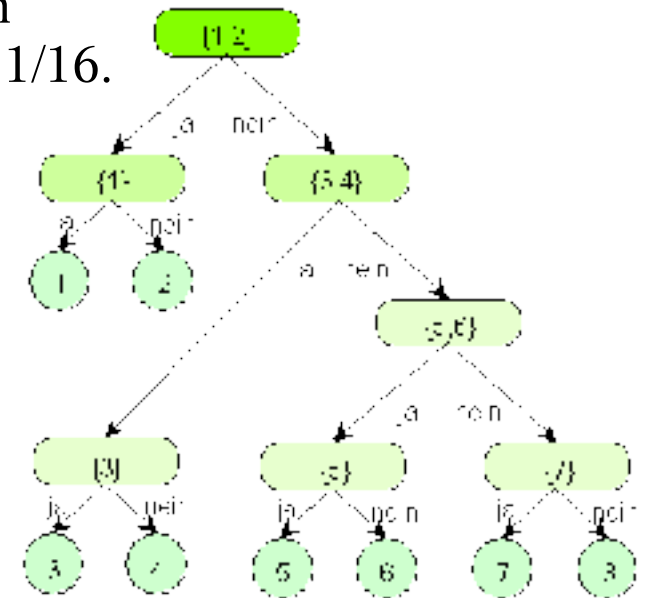
- Im Mittel

$$H = \frac{2}{4} + \frac{2}{4} + \frac{3}{8} + \frac{3}{8} + \frac{4}{16} + \frac{4}{16} + \frac{4}{16} + \frac{4}{16} = 2,75$$

- Formel für die Informationsmenge

$$H = \sum_i -p_i \cdot \log_2 p_i = 2 \cdot \frac{1}{4} \cdot \log_2 \frac{1}{4} + 2 \cdot \frac{1}{8} \cdot \log_2 \frac{1}{8} + 4 \cdot \frac{1}{16} \cdot \log_2 \frac{1}{16} = \frac{2}{4} \cdot 2 + \frac{2}{8} \cdot 3 + \frac{4}{16} \cdot 4 = 2,75$$

- Abfrageschema ist optimal!



Informationstheorie

- Informationsgehalt
 - Information kann gemessen werden
 - alle N Zustände eines Systems gleich wahrscheinlich

$$H = \sum_i -\frac{1}{N} \cdot \log_2 \frac{1}{N} = -N \cdot \frac{1}{N} \cdot \log_2 \frac{1}{N} = \log_2 N$$

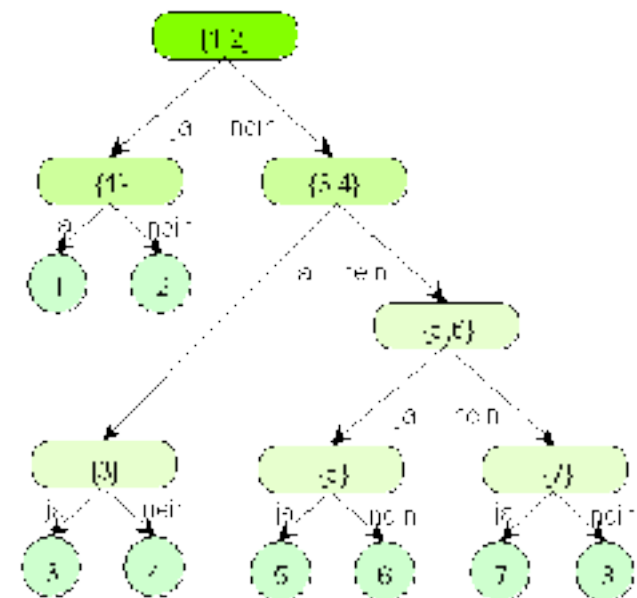
- H heißt *ideelle Entropie*
 - Umfang des Nichtwissens
 - komplementär zum Begriff Informationsmenge
 - H heißt *ideelle Informationsmenge*.
 - Ideelle Entropie
 - Reale, wahre oder wirkliche Entropie
 - Ideelle Entropie immer kleiner oder gleich der realen Entropie

Informationstheorie

- Informationsgehalt
 - wirkliche Informationsmenge i.d.R. größer als ideelle
 - Anzahl (gleichwahrscheinlicher) Zustände Zweierpotenz
 - wirkliche und ideelle Informationsmenge gleich
 - Informationsmenge ist ein Maß
 - Maßeinheit benötigt
 - Bit (b>bit von *binary digit*)
 - Bit ein zweiwertiger (**binär**; *binary*) Zustand
 - jeder Zustand eines Systems mit 2^k Zuständen als Folge von k Bits darstellbar
 - Ja, Nein, Ja, Ja, Nein, Ja, Nein, Nein ($2^8=256$ Zustände)
 - 1 0 1 1 0 1 0 0
- Informatik
 - 8 Bits heißen ein Byte (*byte*)
 - k Bits heißen k -Bitwort
 - Byte ist somit ein 8-Bitwort
 - Einheit für Bit ist b (hier nicht verwendet)
 - Einheit für Byte ist B (hier nicht verwendet)

Informationstheorie

- Huffman-Kodierung (David Huffman)
 - Erstellung eines optimalen Codebaums
 - Zustandswahrscheinlichkeiten aufsteigend ordnen
 - jeweils zwei zusammenzufassen
 - Stets die kleinsten (Summen-)Wahrscheinlichkeiten zusammenfassen
 - 8 7 6 5 4 3 2 1
 - 1/16, 1/16, 1/16, 1/16, 1/8, 1/8, 1/4, 1/4
 - 1/16, 1/16, [1/16, 1/16]=1/8, 1/8, 1/8, 1/4, 1/4
 - 1/8, [1/16, 1/16]=1/8, 1/8, 1/8, 1/4, 1/4
 - 1/8, 1/8, [1/8, 1/8]=1/4, 1/4, 1/4
 - [1/8, 1/8]=1/4, 1/4, 1/4, 1/4
 - 1/4, 1/4, [1/4, 1/4]=1/2
 - [1/4, 1/4]=1/2, 1/2
 - [1/2, 1/2]=1



Informationstheorie

- Huffman-Kodierung (David Huffman)

- Erstellung eines optimalen Codebaums

- Zustandswahrscheinlichkeiten aufsteigend ordnen
- jeweils zwei zusammenzufassen

- Stets die kleinsten (Summen-)Wahrscheinlichkeiten zusammenfassen

- 8_{16} 7_{16} 6_{16} 5_{16} 4_8 3_8 2_4 1_4

- 6_{16} 5_{16} 78_8 4_8 3_8 2_4 1_4

- 56_8 78_8 4_8 3_8 2_4 1_4

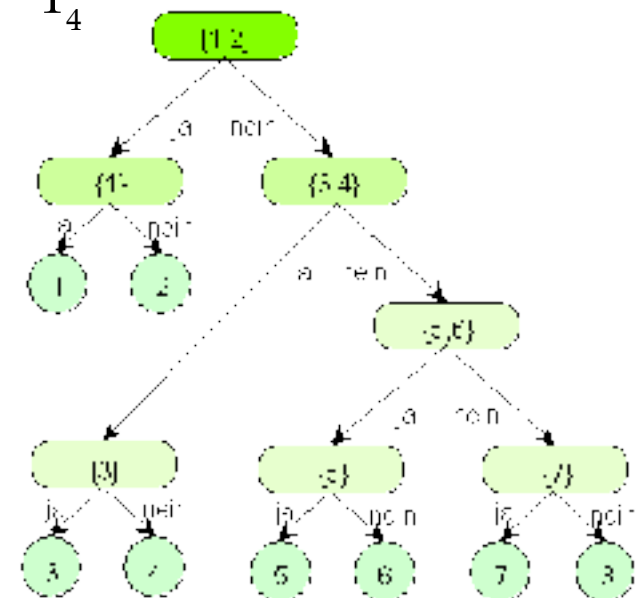
- 3_8 4_8 5678_4 2_4 1_4

- 34_4 5678_4 2_4 1_4

- 2_4 1_4 345678_2

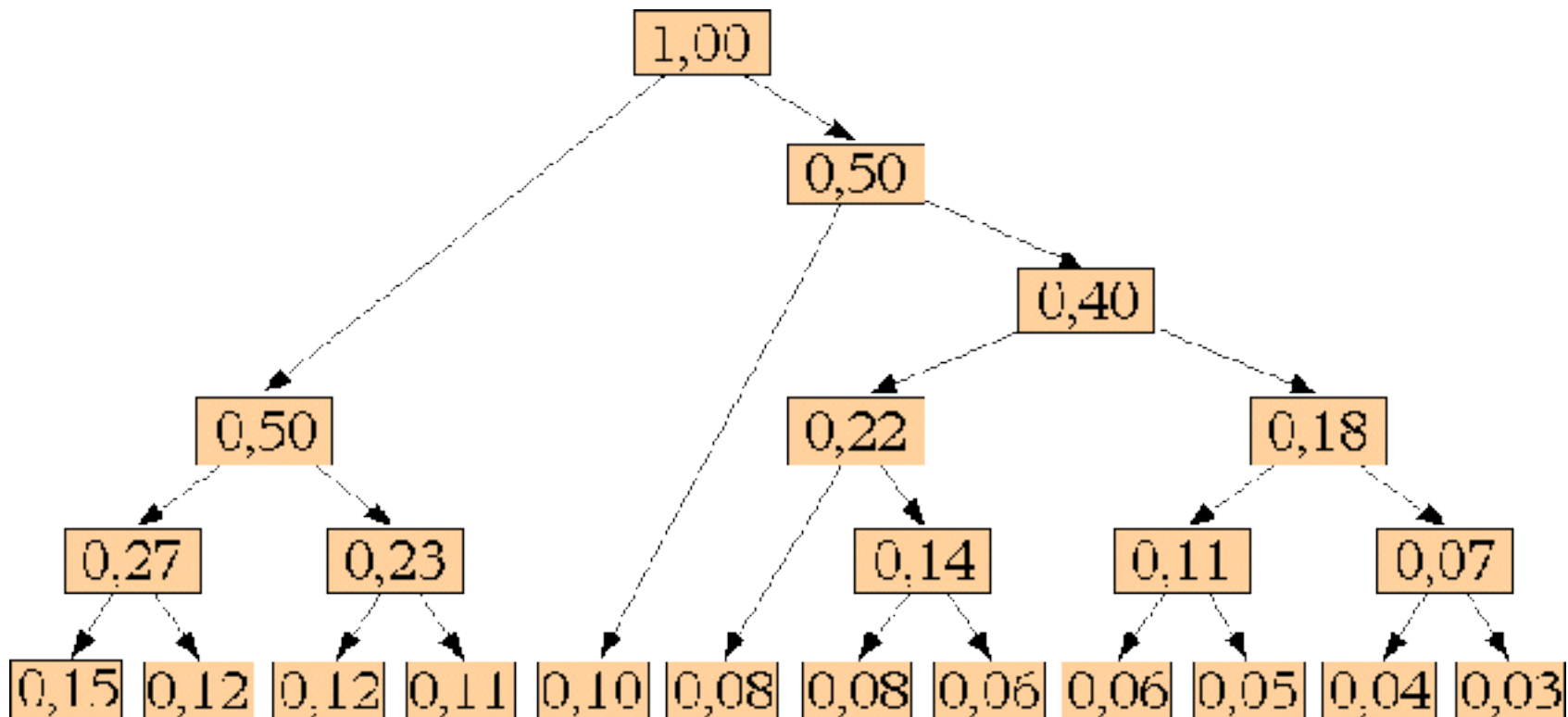
- 345678_2 12_2

- 12345678_1



Informationstheorie


- Huffman-Kodierung (David Huffman)
 - Erstellung eines optimalen Codebaums
 - 12 Elemente mit wahrer Entropie 3,62
 - Es gibt optimaleren Codebaum (s. Aufgabe)



Darstellung von Information

- Darstellung von Information
 - Information ist Wissen
 - braucht Darstellungsform, um über sie sprechen zu können
 - Information nennt man meistens Datum (Mehrz. Daten, data) oder Nachricht (message).
 - synonym
 - Datum
 - statische Repräsentation von Information
 - Nachricht
 - Information während des Transports
 - Definition
 - Wir nennen die statische Repräsentation von Information Datum, die Repräsentation von Information während des Transports Nachricht
 - Datum und Nachricht in Umgangssprache

Darstellung von Information

- Beispiele
 - Das Attribut sei die Farbe eines Eidotters.
 - Die Ausprägungen dieses Attributs können u.a. folgendermaßen dargestellt werden
 - 0, 1, 2, 3, 4, 5, 6
 - A, B, C, D, E, F, G
 - I, II, III, IV, V, VI, VII
 - weiß, blassgelb, hellgelb, gelb, orange, rot, dunkelrot
 -  , , , , , , ,
 - Darstellungsform von Information durch Daten hängt von beabsichtigter Verwendung
 - Information von Menschen verarbeitet
 - Daten einfach darstellbar
 - Daten einfach verstehbar
 - markante Unterschiede zwischen Daten
 - früher die einzige Form der Informationsverarbeitung

Darstellung von Information

- Beispiel
 - Die ägyptischen Hieroglyphen (griech.: „heilige Eingrabungen“)
 - zunächst nur für in Stein gekerbten Schriften
 - schön anzusehen
 - Bedeutung aus konsonantischen Lautwert erschließen
 - Piktogramme schwierig anzufertigen
 - vereinfachte Schreibschrift entwickelt → Demotische
 - einfacher zu schreibenschwieriger zu lesen

Darstellung von Information



Darstellung von Information

- Mensch verwendet für Darstellung von Information
 - Schriftzeichen (s.a. <http://www.unicode.org>)
 - lateinische Alphabet
 - arabische Schrift
 - griechische: α , β , γ , ...
 - chinesische
 - japanische
 - spezielle indische Alphabete
 - Alphabete
 - einfach aufschreiben
 - nach menschlichen Maßstäben leicht lesbar

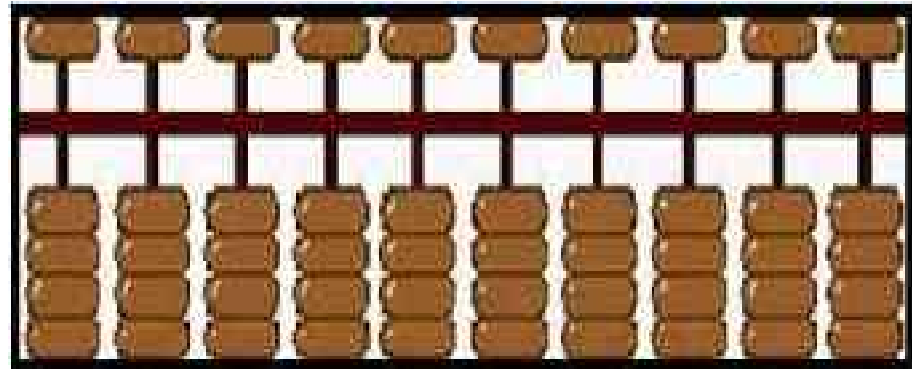
Darstellung von Information

- Maschinen Information verarbeiten

- Daten maschinenlesbar sein
- verschiedene Mechanismen entwickelt

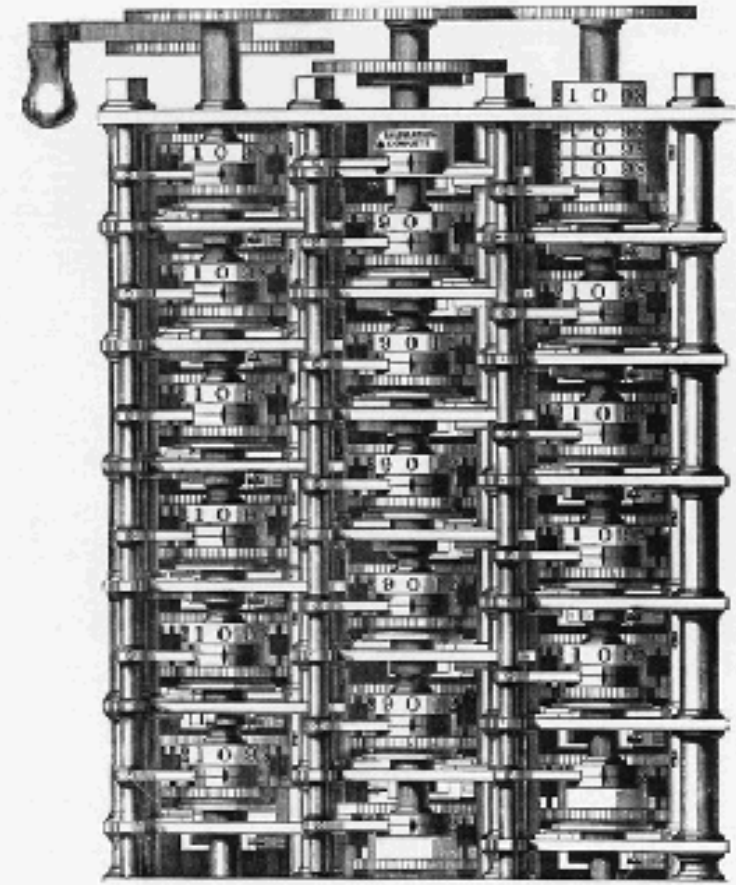
- Beispiel

- Darstellung von Daten im Abakus
- auf Stange (Setzkasten) Steinchen oder Kugeln
- Anzahl repräsentiert die Zahl
- komplexe Berechnungen durch geschicktes Umschichten oder Verschieben
- Abakus noch heute in Gebrauch.



Darstellung von Information

- Zahlwerten durch Stellung von Zahnrädern
 - in Rechenmaschinen verwendet
 - Pascal, Leibniz
 - moderne Rechenmaschine
 - Charles Babbages
 - Differential Engine und die Analytical Engine
 - Differential-Analysator von V. Bush (ca. 1930)



Engraving from a woodcut of a small portion of Mr. Babbage's Difference Engine No. 1, the property of Government, at present deposited in the Museum at South Kensington.

It was constructed 1833.
This portion put together 1848.
The construction abandoned 1848.
This plate was printed June, 1853.
This portion was in the Exhibition 1854.

Darstellung von Information

- Konrad Zuses Rechenmaschine
 - elektromagnetische Schalter, sog. Relais
 - Lochbänder
 - mechanische Webstuhl von Jacquard
 - Lochkarten
 - Maschinen von Hollerith
 - durch geeignete Geräte direkt weiter verarbeitbar
- geschriebene Schrift nur schwierig maschinell behandelbar
- Moderne Rechenmaschinen
 - Repräsentation von Information in Form elektrischer Ladungen
 - Zustand eines Transistors
 - magnetische Feldlinien, (Diskette, Magnetband)
- optische Systeme
 - wiederbeschreibbare CDs

Darstellung von Information

- Unterschiedliche Repräsentation
 - vom Menschen lesbare Daten
 - maschinenlesbare Daten
 - beide angleichen
 - in mechanischen Rechenmaschinen Ziffern auf Walzen angebracht
- Darstellung von Daten in Rechensystemen
 - in festen Einheiten
 - Wörter (word)
 - Länge kann von Maschine zu Maschine unterschiedlich sein
 - Vergangenheit sehr unterschiedliche Wortlängen
 - 1 Bit
 - 16 Bits
 - 32 Bits
 - 36 Bits
 - 64 Bits und länger

Darstellung von Information

- heute Wörter in Rechenmaschinen in
 - Bytes
 - jedes Byte (d.h. 8 Bits bzw. 256 Werte)
 - einzeln adressiert
 - Speichergröße in Bytes (kBytes, MBytes, GBytes)
 - Adresslänge voneinander abhängig
 - ältere Mikrorechnerarchitekturen
 - nur $2^{16}=65.536$ Bytes Speicherstellen
 - nach heutigen Maßstäben sehr kleiner Adressraum
 - nur durch gewisse Tricks umgehbar
 - heute mehr als 2^{24} Bytes adressieren
 - Moderne Entwicklungen: 2^{64} Bytes adressierbar

Darstellung von Information

- Rechnersprachen
 - mehrere Bytes als ein Objekt betrachtet
 - kurze Zahlen (*short integer*: 1 Byte)
 - Zahlen (*integer*: 2 Bytes oder 4 Bytes)
 - lange Zahlen (*long integer*: 4 Bytes oder 8 Bytes)
 - genaue Anzahl von Bytes in Sprachbeschreibungen
 - in Java
 - **byte** (8 Bits)
 - **short** (16 Bits)
 - **int** (32 Bits)
 - **long** (64 Bits)
 - Daten in Bytes oder Vielfachen von Bytes dargestellt

Darstellung von Information

- Rechnersprachen
 - Daten in Bytes oder Vielfachen dargestellt
 - Kodierung (coding)
 - Abbildung zwischen Datum und Darstellung
 - auch kürzer Code
 - Beispiele
 - Flagencode
 - Morsecode
 - häufige Buchstaben kürzeste Zeichen
 - Zwischen zwei Zeichen längere Pause

a	· -	b	- · · ·	c	- · · ·	d	- · ·	e	·	f	· · · ·
g	- · ·	h	· · · ·	i	· ·	j	· - - -	k	- · -	l	· · · ·
m	- -	n	- ·	o	- - -	p	· - - ·	q	- · - -	r	· - ·
s	· · ·	t	-	u	· · -	v	· · · -	w	· - -	x	- · · -
y	- · - -	z	- · · ·	ch	- - - -	ä	· - - -	ö	- - - ·	ü	· · - -

Darstellung von Information

- Endliche Menge von Symbolen heißt Zeichenvorrat
 - Symbole nennt man Zeichen
 - Beispiele Zeichenvorrat
 - lateinischen Buchstaben
 - die Ziffern
 - Symbole für Farbe von Eidottern: , ● , ● , ● , ● , ● , ●
- Geordneter Zeichenvorrat heißt Alphabet
 - Alphabet allgemeiner als umgangssprachlich
 - ASCII-Alphabet
 - EBCDIC
 - Unicode
 - Code ist Zuordnung zwischen zwei Zeichenvorräten
 - Erstellen eines Codes wird als Kodierung bezeichnet

Darstellung von Information

- Fano-Bedingung
 - jedes kürzere Wort eindeutig von längeren unterscheidbar
 - alle Wörter haben unterschiedliche Anfangsfolgen
- Codes können eindeutig oder mehrdeutig sein
 - Kodierungen können injektiv und bijektiv sein
 - Codes
 - Darstellung eines Datums
 - Text von einem Papier in Textverarbeitungssystem
 - gedruckte Zeichen in internes Bitwort abbilden
 - Zahlen
 - Bilder
 - akustische Signale
 - Kryptographie (griech.: geheimes Schreiben)
 - Chiffrieren (encipher),
 - Dekodierung oder Dechiffrieren (decipher)

Darstellung von Information

- Kodierungen
 - bei Übertragung von Information
 - möglichst geringe Zeichenlängen
 - Grenzen durch Informationstheorie
 - Fehler sollen entdeckt oder korrigiert werden
 - optimale Kodierung zu aufwendig
 - Nachrichten (transportierte Information) und Daten unterschiedliche Darstellung
 - andere physikalische Medien
 - elektromagnetische Wellen (Funk)
 - Lichtwellen (Lichtwellenleiter: LWL; Glasfaser fibre optics)
 - akustische Signale (Akustikkoppler)
 - Leitungskodierung
 - in Nachrichtentechnik behandelt
 - Datum und Nachricht nicht immer exakt voneinander abgrenzbar