

# Kanäle, Codecs und Medien

Jörg Cassens

Institut für Mathematik und Angewandte Informatik

Medieninformatik  
WS 2017/2018



medieninformatik

IMA – Institut für  
Mathematik und  
Angewandte Informatik



 [pingo.upb.de/736695](https://pingo.upb.de/736695)

## 1 Kanäle und Medien

## 2 Digitalisierung

## 3 Kompression

- Klares Verständnis der grundlegenden Gesetzmäßigkeiten bei der Digitalisierung
- Welche Probleme gibt es bei der
  - Codierung
  - Kompressionvon Mediendaten?
- Einführung universeller Kompressionsverfahren

# Charakterisierung medialer Angebote

## 1 Sehen

- Kulturell weit entwickelt bis zum Lesen von Texten
- Visuelle Erscheinung zentrales Element vieler Medienangebote

## 2 Hören

- Eng mit Umweltwahrnehmung verbunden, wenig bewußt

## 3 Riechen

- Sehr wenig bewußt, beeinflusst Stimmungen

## 4 Schmecken

- Hoch spezialisiert, wenig zur Übermittlung von Informationen geeignet

## 5 Tasten

- Verschiedene Untergruppen (Druck, Berührung, Temperatur)

Bei Computern in der Praxis derzeit zumeist 1 und 2, vereinzelt 5.

## Definition

“**Multimedia** ist der Trend, die verschiedenen Kommunikationskanäle des Menschen mit den Mittel der Informationswissenschaft über alle Quellen [Modalitäten] zu integrieren und als Gesamtheit für die Kommunikation zu nutzen”

Peter Henning (2003)

## Definition

“**Multimedia** ist der Trend, die verschiedenen **Kommunikationskanäle** des Menschen mit den Mittel der Informationswissenschaft über alle **Quellen [Modalitäten]** zu integrieren und als Gesamtheit für die Kommunikation zu nutzen”

Peter Henning (2003)

Wir können Medien auf verschiedenen Ebenen unterscheiden:

- **Präsentation** & Aufnahme
  - Die “technische Seite”
  - Hilfsmittel zur Ein- und Ausgabe von Informationen
  - Welche Geräte werden gebraucht (Mikrofon, Kamera, Lautsprecher)
- **Codierung**
  - Die Repräsentation
  - In welcher Form wird die Information beschrieben
- **Wahrnehmung** & Produktion
  - Die “menschliche Seite”
  - Der genutzte/angesprochene Sinneskanal

## ■ **Präsentation – Medial**

- Radio: monomedial
- TV: multimedial

## ■ **Codierung – Codal**

- Nur Text, nur Graphik: monocodal
- Gemischt: multicodal

## ■ **Wahrnehmung – Modal**

- Nur die Augen ansprechend: monomodal
- Augen und Ohren ansprechend: multimodal

Problem: Unterschiedliche Belegung in unterschiedlichen Kontexten und Fachgebieten

## Video 5.1: Modality, Codality, Mediality

Kanäle und  
Medien

Digitalisierung

Kompression



👉 Kiss machine – The Big Bang Theory (1:33)

## 1 Kanäle und Medien

## 2 Digitalisierung

- Analog und Digital
- Abtastung
- Codierung

## 3 Kompression

# Analog und Digital

## Definition

**Analoges Signal:** Deterministische und kontinuierliche Änderung einer physikalischen Größe entsprechend eines Meßwertes der zu übertragenden Größe.

## Definition

**Analoges Signal:** Deterministische und kontinuierliche Änderung einer physikalischen Größe entsprechend eines Meßwertes der zu übertragenden Größe.

## Definition

**Digitales Signal:** Annäherung an den Meßwert in einem festen Raster über die Zeit oder den Raum mittels eines endlichen Vorrats möglicher Repräsentationen.

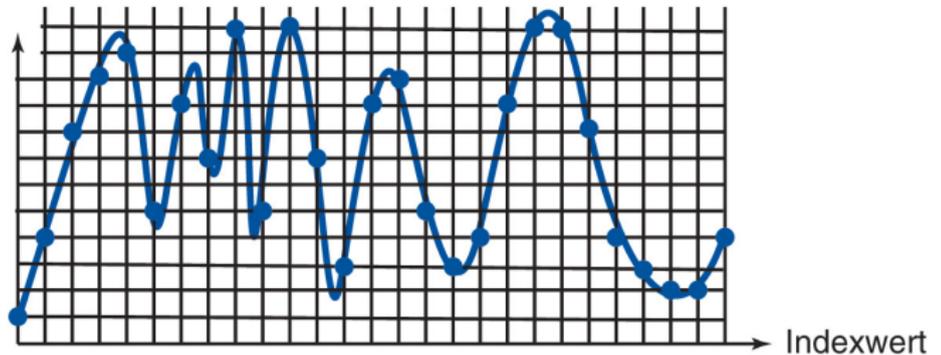
## Definition

**Analoges Signal:** Deterministische und **kontinuierliche** Änderung einer physikalischen Größe entsprechend eines Meßwertes der zu übertragenden Größe.

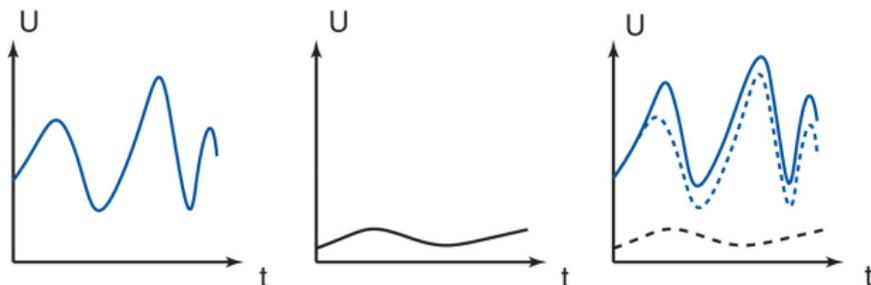
## Definition

**Digitales Signal:** Annäherung an den Meßwert in einem **festen Raster** über die Zeit oder den Raum mittels eines **endlichen Vorrats** möglicher Repräsentationen.

Signalwert



**Abbildung 2.1:** Grundprinzip der Digitalisierung



**Abbildung 2.2:** Analoges Signal und Störsignal: links Originalsignal, Mitte Störsignal, rechts resultierendes analoges Signal

# Signale und Störungen

Kanäle und Medien

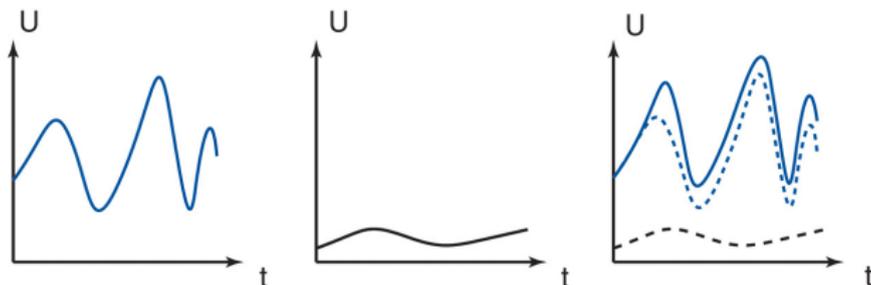
Digitalisierung

Analog und Digital

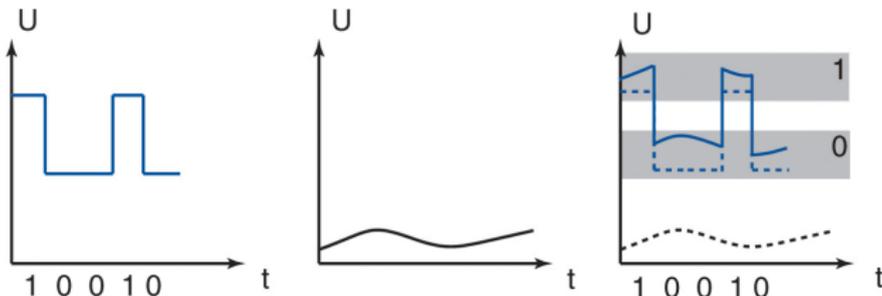
Abtastung

Codierung

Kompression



**Abbildung 2.2:** Analoges Signal und Störsignal: links Originalsignal, Mitte Störsignal, rechts resultierendes analoges Signal



**Abbildung 2.3:** Digitales Signal und Störsignal: links Originalsignal, Mitte Störsignal, rechts resultierendes analoges Signal

# Abtastung

- Jedes Signal, welches am Ende der Verarbeitungskette als mediales Angebot präsentiert wird, muß irgendwann in digitaler Form erscheinen
  - Synthese: das Ausgangsmaterial wird in digitaler Form hergestellt
  - Abtastung: ein analoges Signal wird in ein digitales umgewandelt
- Abtastung:
  - Wie *häufig*? → Sampling, Diskretisierung
  - Wie *genau*? → Quantisierung

## Definition

**Diskretisierung:** Festes Raster von Meßpunkten gleichen Abstands auf der Achse oder den Achsen, auf der sich das analoge Signal verändert. Die Dichte wird als Abtastrate bezeichnet. Zu jedem Meßpunkt wird jeweils das Signal (Sample) bestimmt.

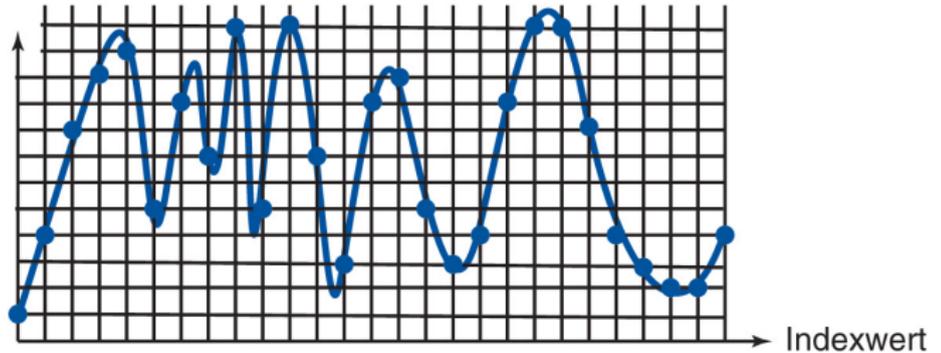
## Definition

**Diskretisierung:** Festes Raster von Meßpunkten gleichen Abstands auf der Achse oder den Achsen, auf der sich das analoge Signal verändert. Die Dichte wird als Abtastrate bezeichnet. Zu jedem Meßpunkt wird jeweils das Signal (Sample) bestimmt.

## Definition

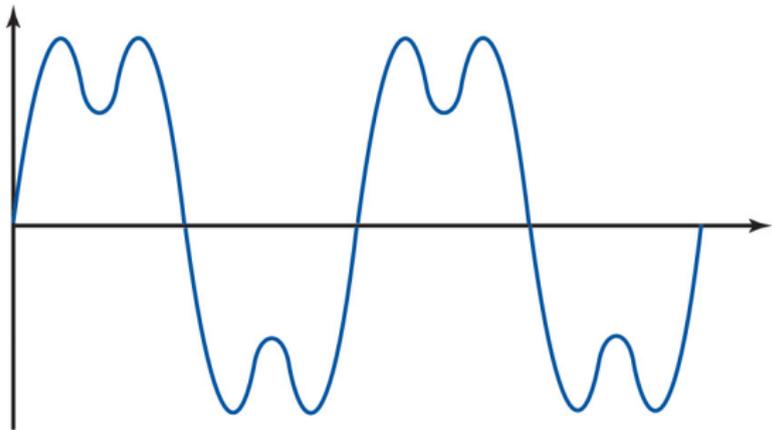
**Quantisierung:** Darstellung der ermittelten Meßwerte in einem festen Werteraster. Die Zahl der zur Verfügung stehenden Zahl der Bits wird auch als Auflösung der Digitalisierung bezeichnet.

Signalwert

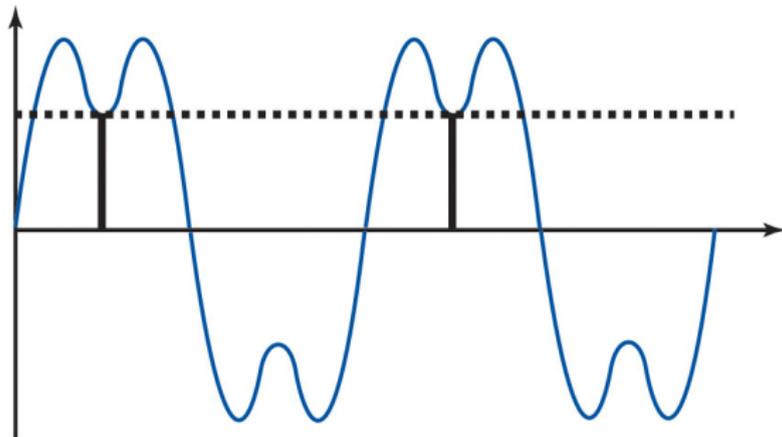


**Abbildung 2.1:** Grundprinzip der Digitalisierung

- Man beachte die Abweichungen vom Originalsignal



**Abbildung 2.4:** Ein periodisches Beispielsignal



**Abbildung 2.5:** Abtastung mit zu niedriger Abtastrate I

# Abtasttheorem III

Kanäle und  
Medien

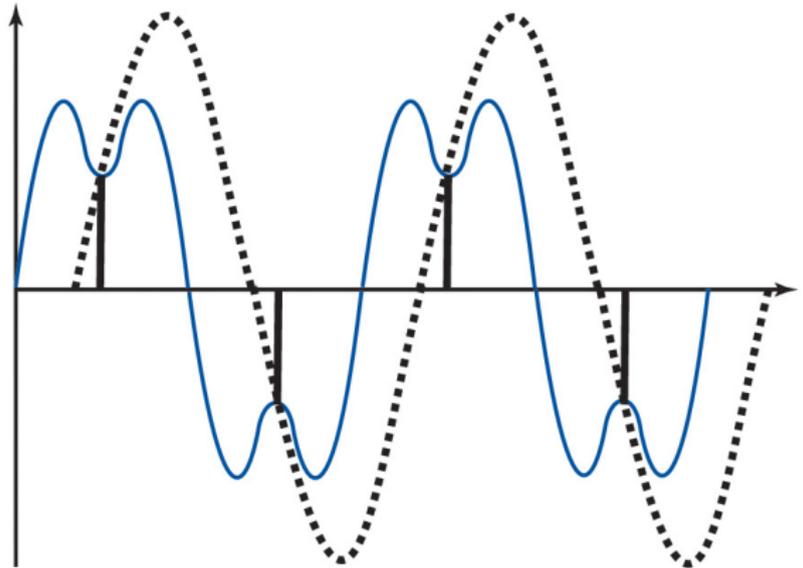
Digitalisierung

Analog und Digital

**Abtastung**

Codierung

Kompression



**Abbildung 2.6:** Abtastung mit zu niedriger Abtastrate II

# Abtasttheorem IV

- 1 Ein reales Signal kann als Überlagerung verschiedener Grundsignale verschiedener Frequenzen aufgefaßt werden.
- 2 Für die Wahl der Abtastrate ist die Frequenz des im digitalen Signal enthaltenen Anteils mit der höchsten Frequenz entscheidend.

## Definition

**Abtasttheorem** (Shannon, Nyquist, Whittaker, Kotelnikow):  
Wenn ein kontinuierliches Signal mit einer oberen Grenzfrequenz  $f_{max}$  mit einer Abtastrate von mehr als  $2 * f_{max}$  abgetastet wird, kann man das Ursprungssignal ohne Informationsverlust aus dem abgetasteten Signal rekonstruieren.

- Jean Baptiste Joseph Fourier (1768-1830)
- Jede periodische Schwingung kann durch eine (unendliche) Summe von überlagerten Cosinus-Schwingungen angenähert werden
- Überlagerung harmonischer Schwingungen

$$x(t) = \sum_{k=0}^{\infty} a_k * \cos(k\omega_0 t + \theta_k)$$

# Beispiel

Kanäle und  
Medien

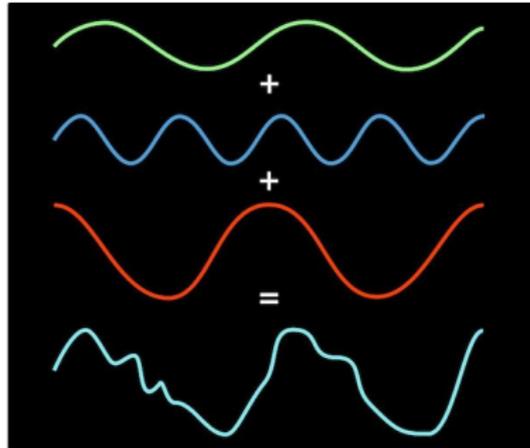
Digitalisierung

Analog und Digital

Abtastung

Codierung

Kompression



Christine Daniloff/MIT

Kanäle und  
Medien

Digitalisierung

Analog und Digital

**Abtastung**

Codierung

Kompression

👉 Lucas V. Barbosa

# Abtasttheorem Revisited

Kanäle und Medien

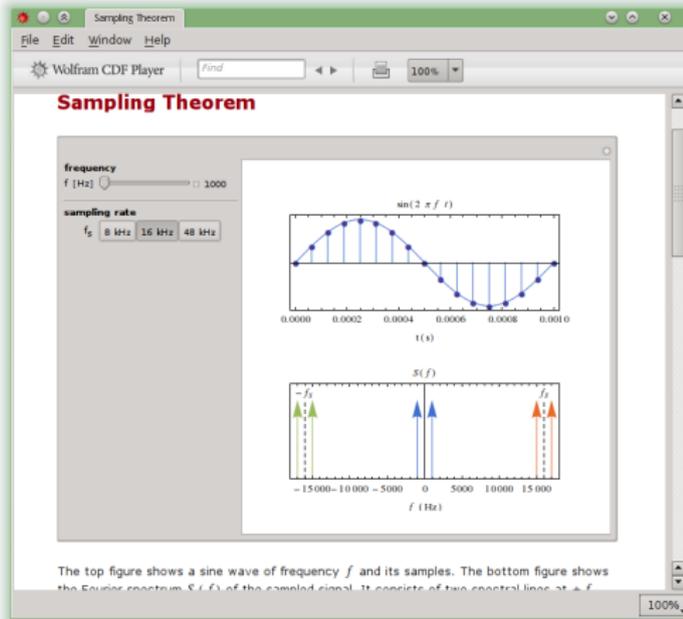
Digitalisierung

Analog und Digital

Abtastung

Codierung

Kompression



[demonstrations.wolfram.com/SamplingTheorem/](https://demonstrations.wolfram.com/SamplingTheorem/)

- Hohe Frequenzanteile enthalten häufig nur geringe Energie (Oberschwingungen bei Musikinstrumenten)
- Noch wichtiger: Begrenzung der menschlichen Sinne
- Daher: Bandbreitenbeschränkung, danach Abtastung mit dem doppelten der höchsten verbleibenden Frequenz
  - Beispiel Bandbreitenbeschränkung auf 5 bis 22000 Hz bei Audiosignalen auf einer CD

- Bei der Rekonstruktion können Signalanteile entstehen, die im Originalsignal nicht vorhanden waren
- Entsteht z.B. durch eine zu kleine Abtastrate
- Eine theoretisch mögliche vollständige Rekonstruktion kann durch Eigenschaften des Ausgabegerätes zu Aliasing-Effekten führen (Moiré-Effekt bei Graphik auf kleiner Auflösung)
- Abhilfe: Ausfiltern zu hoher Frequenzanteile (“Weichzeichner”)

# Codierung

- Information ist ein nichtstoffliches Phänomen, das durch die Interpretation bestimmter Zeichen oder Zeichenfolgen entsteht, die man die Repräsentation der betreffenden Information nennt.
- Beispiel Uhrzeit: Ziffernfolge eine Digitaluhr, analoge Uhr mit Zeigern, sprachliche Zeichen, Mengentheorie
- Informationstheorie nach Shannon analysiert Codierungen vor allem unter dem Aspekt des Auftretens eines bestimmten Zeichen eines Zeichenvorrats (stochastischer Ansatz)
- Grundlegend: Nachricht und Nachrichtenquelle.

- Wir nehmen Texte als Beispiel
- Ein Bild oder Musikstück liegt aber ebenfalls als lange Folge von synthetisch erzeugten oder abgetasteten Zeichen vor
- Daher sind die folgenden Überlegungen für alle Medientypen relevant
- In der Informatik ist besonders die Repräsentation im Binärformat relevant, also als Abfolge von 0 und 1

## Ausgangspunkt

Wir haben eine Codierung von Informationen in einem beliebigen Ausgangsformat und suchen eine möglichst effiziente Codierung im Binärformat

## Definition

**Zeichenvorrat:** Endliche Menge von Zeichen.

## Definition

**Zeichenvorrat:** Endliche Menge von Zeichen.

**Nachricht:** Eine Nachricht im Zeichenvorrat  $A$  ist eine endliche Sequenz von Zeichen  $a$  aus  $A$ .

## Definition

**Zeichenvorrat:** Endliche Menge von Zeichen.

**Nachricht:** Eine Nachricht im Zeichenvorrat  $A$  ist eine endliche Sequenz von Zeichen  $a$  aus  $A$ .

**Codierung:** Seien  $A$  und  $B$  Zeichenvorräte. Dann ist eine Codierung  $c$  von  $A$  in  $B$  eine Abbildung von Nachrichten aus  $A$  in Nachrichten aus  $B$ :  $c(a) \rightarrow b, a \in A, b \in B$ .

## Definition

**Zeichenvorrat:** Endliche Menge von Zeichen.

**Nachricht:** Eine Nachricht im Zeichenvorrat  $A$  ist eine endliche Sequenz von Zeichen  $a$  aus  $A$ .

**Codierung:** Seien  $A$  und  $B$  Zeichenvorräte. Dann ist eine Codierung  $c$  von  $A$  in  $B$  eine Abbildung von Nachrichten aus  $A$  in Nachrichten aus  $B$ :  $c(a) \rightarrow b, a \in A, b \in B$ .

Im folgenden betrachten wir nur die Einzelzeichencodierung.

## Definition

**Zeichenvorrat:** Endliche Menge von Zeichen.

**Nachricht:** Eine Nachricht im Zeichenvorrat  $A$  ist eine endliche Sequenz von Zeichen  $a$  aus  $A$ .

**Codierung:** Seien  $A$  und  $B$  Zeichenvorräte. Dann ist eine Codierung  $c$  von  $A$  in  $B$  eine Abbildung von Nachrichten aus  $A$  in Nachrichten aus  $B$ :  $c(a) \rightarrow b, a \in A, b \in B$ .

Im folgenden betrachten wir nur die Einzelzeichencodierung.

**Nachrichtenquelle:** Eine Nachrichtenquelle nach Shannon ist ein Zeichenvorrat  $A$  zusammen mit einer Wahrscheinlichkeitsverteilung, die für jedes Zeichen  $a \in A$  die Wahrscheinlichkeit seines Auftretens angibt.

Wahrscheinlichkeit von  $a \in A$  wird mit  $p_a$  bezeichnet,  
 $0 \leq p_a \leq 1, \sum p_a = 1$

# Beispiel Nachrichtenquelle I

Zeichen $a$	A	B	C	D
Wahrscheinlichkeit $p_a$ in Quelle 1	1.0	0.0	0.0	0.0
Wahrscheinlichkeit $p_a$ in Quelle 2	0.25	0.25	0.25	0.25
Wahrscheinlichkeit $p_a$ in Quelle 3	0.5	0.25	0.125	0.125

Tabelle 2.1: Drei Beispiele für Nachrichtenquellen

- Unterschiedlicher Informationsgehalt der einzelnen Zeichen
  - Quelle 1: Nur A, bringt keine weiteren Informationen
  - Quelle 2: Jedes Zeichen bringt gleichen Informationsgehalt entsprechend der Auswahl eines Elements aus einer vierelementigen Menge
  - Quelle 3: A bringt weniger Informationen, es ist “weniger überraschend”
- Informationsgehalt beruht ausschließlich auf der Wahrscheinlichkeit des Auftretens, keine andere Verständnisebene notwendig

# Beispiel Nachrichtenquelle II

Kanäle und  
Medien

Digitalisierung

Analog und Digital

Abtastung

Codierung

Kompression

- Wie kann man den Informationsgehalt messen?
- 1 Bit ist die Informationsmenge, die notwendig ist, um aus zwei Elementen eines auszuwählen (Entscheidungsgehalt)
  - 2 Bit = vier Elemente, 3 Bit = acht Elemente, ...

Zeichen $\alpha$	A	B	C	D
Entscheidungsgehalt in Quelle 1 [Bit]	0	undefiniert	undefiniert	undefiniert
Entscheidungsgehalt in Quelle 2 [Bit]	2	2	2	2
Entscheidungsgehalt in Quelle 3 [Bit]	1	2	3	3

Tabelle 2.2: Beispiele für den Informationsgehalt von Zeichen

# Auftrittswahrscheinlichkeit und Bit

Kanäle und  
Medien

Digitalisierung

Analog und Digital

Abtastung

Codierung

Kompression

- Gesucht: die Zahl, die zur Basis 2 angewendet die Zahl der Auswahlmöglichkeiten ergibt
- Zahl der Auswahlmöglichkeit Kehrwert der Wahrscheinlichkeit
- Umrechnung der Auftrittswahrscheinlichkeit in den Informationsgehalt in Bit:

$$x_a = \log_2 \left( \frac{1}{p_a} \right)$$

- Stochastische Betrachtung, also “krumme” Bitwerte möglich ( $p = 0.3, 1.74$  Bit)

- Entropie ist das “Maß der Unordnung”
- Quelle 2 ist chaotisch, alle Zeichen treten gleich häufig auf
- Quelle 3 gibt hingegen das A bevorzugt aus

## Definition

Die **Entropie** einer Nachrichtenquelle ist der durchschnittliche Entscheidungsgehalt eines Zeichens der Nachrichtenquelle:

$$H = \sum_{a \in A} p_a * x_a = \sum_{a \in A} p_a * \log_2 \left( \frac{1}{p_a} \right)$$

# Beispiel revisited

Kanäle und  
Medien

Digitalisierung

Analog und Digital

Abtastung

Codierung

Kompression

Zeichen $a$	A	B	C	D
Wahrscheinlichkeit $p_a$ in Quelle 1	1.0	0.0	0.0	0.0
Wahrscheinlichkeit $p_a$ in Quelle 2	0.25	0.25	0.25	0.25
Wahrscheinlichkeit $p_a$ in Quelle 3	0.5	0.25	0.125	0.125

Tabelle 2.1: Drei Beispiele für Nachrichtenquellen

- Quelle 1: Entropie von 0
- Quelle 2: Entropie von 2
- Quelle 3: Entropie von 1,75

## Definition

**Wortlänge:** Die Menge der Wörter aus einem Zeichenvorrat  $A$  wird mit  $A^*$  bezeichnet. Für ein Wort  $w \in A^*$  ist die Länge des Wortes die darin enthaltene Anzahl von Zeichen, bezeichnet mit  $|w|$ . Wenn eine Codierung  $c$  einem Zeichen  $a \in A$  ein Wort  $c(a) \in B^*$  zuweist, dann ist  $|c(a)|$  die Wortlänge der Codierung von  $a$ .

- In einem Wort können Zeichen auch mehrfach auftreten

- Wir sind in erster Linie an binären Codierungen interessiert
- Wenn wir wissen wollen, wieviel Platz die Codierung eines einzelnen Zeichens im Durchschnitt belegt, können alle möglichen Codierungen unter Einbeziehung der Auftrittswahrscheinlichkeit eines Zeichens betrachten

## Definition

Die **durchschnittliche Wortlänge**  $L$  ist die nach Auftrittswahrscheinlichkeiten gewichtete Summe der Wortlängen aller Codierungen:

$$L = \sum_{a \in A} p_a * |c(a)|$$

# Beispiel

Kanäle und  
Medien

Digitalisierung

Analog und Digital

Abtastung

Codierung

Kompression

<b>Zeichen <math>a</math></b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
Codierung $c_1$	<b>00</b>	<b>01</b>	<b>10</b>	<b>11</b>
Codierung $c_2$	<b>0</b>	<b>10</b>	<b>110</b>	<b>111</b>

Tabelle 2.3: Zwei Beispiele für Codierungen

# Beispiel: Codierung 1

Kanäle und  
Medien

Digitalisierung

Analog und Digital

Abtastung

Codierung

Kompression

Zeichen $a$	A	B	C	D
Wahrscheinlichkeit $p_a$ in Quelle 3	0.5	0.25	0.125	0.125
Codierung $c_1$	<b>00</b>	<b>01</b>	<b>10</b>	<b>11</b>
Wortlänge	2	2	2	2
Durchschnittliche Wortlänge $L = 0.5 \cdot 2 + 0.25 \cdot 2 + 0.125 \cdot 2 + 0.125 \cdot 2 = 2$				

Tabelle 2.4: Beispiel für eine redundante Codierung

- Die mittlere Wortlänge im Beispiel beträgt 2
- Die Entropie liegt wie vorher gesehen bei 1.75

# Beispiel: Codierung 2

Kanäle und Medien

Digitalisierung

Analog und Digital

Abtastung

Codierung

Kompression

Zeichen $a$	A	B	C	D
Wahrscheinlichkeit $p_a$ in Quelle 3	0.5	0.25	0.125	0.125
Codierung $c_2$	<b>0</b>	<b>10</b>	<b>110</b>	<b>111</b>
Wortlänge	1	2	3	3
Durchschnittliche Wortlänge $L = 0.5 \cdot 1 + 0.25 \cdot 2 + 0.125 \cdot 3 + 0.125 \cdot 3 = 1.75$				

Tabelle 2.5: Beispiel für eine optimale Codierung

- Die mittlere Wortlänge im Beispiel beträgt 1.75
- Die Entropie liegt wie vorher gesehen bei 1.75

## Definition

**Redundanz:** Die Redundanz  $R$  einer binären Codierung für eine Informationsquelle ist die Differenz der mittleren Wortlänge und der Entropie:

$$R = L - H$$

- Warum können  $L$  und  $H$  in einer Formel auftauchen?

## Definition

**Redundanz:** Die Redundanz  $R$  einer binären Codierung für eine Informationsquelle ist die Differenz der mittleren Wortlänge und der Entropie:

$$R = L - H$$

- Warum können  $L$  und  $H$  in einer Formel auftauchen?
- Wir können das machen, weil die Wortlänge bei Binärformaten die Größe in Bit angibt, dieser Wert also die gleiche Einheit hat wie die Entropie

Zeichen $a$	A	B	C	D
Wahrscheinlichkeit $p_a$ in Quelle 3	0.5	0.25	0.125	0.125
Codierung $c_2$	<b>0</b>	<b>10</b>	<b>110</b>	<b>111</b>
Wortlänge	1	2	3	3
Durchschnittliche Wortlänge $L = 0.5 \cdot 1 + 0.25 \cdot 2 + 0.125 \cdot 3 + 0.125 \cdot 3 = 1.75$				

Tabelle 2.5: Beispiel für eine optimale Codierung

## Definition

Eine **Codierung** einer Nachrichtenquelle heißt **optimal**, wenn die Redundanz der Codierung gleich Null ist.

# Outline

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

1 Kanäle und Medien

2 Digitalisierung

3 Kompression

- Stochastische Codierung
- Andere Verfahren

## Definition

Eine **Kompressionsverfahren**, das für alle Daten unabhängig von Ursprung und Bedeutung angewendet werden kann heißt **universell**. Verfahren, die nur auf Daten eines bestimmten Typs (Audio, Bild) anwendbar sind heißen **speziell**.

- Universell: zip, gzip, bzip2
- Speziell: z.B. Ausnutzen von Eigenschaften, wie die daß es in Photos häufig Flächen gleichen Farbtons gibt.

## Definition

Eine **Kompressionsverfahren** heißt **verlustbehaftet**, wenn bei der Kompression Informationen aus dem Original verloren gehen. Im Gegensatz stehen **verlustfreie** Methoden, die die vollständige und genaue Rekonstruktion des Originals ermöglichen.

	Universelle Verfahren	Spezielle Verfahren
Verlustfreie Verfahren	Beispiele: Huffman, LZW	Beispiele: PNG, AIFF
Verlustbehaftete Verfahren	(nicht sinnvoll)	Beispiele: JPEG, MP3

Tabelle 2.6: Klassifikation von Kompressionsverfahren

# Stochastische Codierung

- Universelle, verlustfreie Kompression
- Vertreter einer Gruppe von Verfahren, die auf Shannon's informationstheoretischen Überlegungen basiert
- Statistische oder stochastische Verfahren
- Ausgangspunkt: Shannonsche Informationsquelle mit einer Angabe von Auftretswahrscheinlichkeiten
- Der Huffman Algorithmus konstruiert zu einer Nachrichtenquelle eine Codierung, die unter bestimmten Bedingungen optimal ist (wie beim Morse-Code: e=".", t="-", f="..-", q="--.-")

# Huffman-Codierung II

- Problem: variable Codelänge
- Lösungsmöglichkeit
  - Trennzeichen
  - Aufbau (Morse Code: Pausen)

- Problem: variable Codelänge
- Lösungsmöglichkeit
  - Trennzeichen
  - Aufbau (Morse Code: Pausen)
- Hier: Fano-Bedingung

## Definition

**Fano-Bedingung:** Eine Codierung  $C$  eines Zeichenvorrats  $A$  in einen Zeichenvorrat  $B$  erfüllt die Fano-Bedingung, wenn für alle Zeichen  $x$  und  $y$  aus  $A$  gilt, daß das Wort  $c(x)$  nicht Anfang des Wortes  $c(y)$  ist, also keiner der verwendeten Codes Anfang eines anderen ist.

Zeichen $a$	A	B	C	D
Codierung $c_1$	00	01	10	11
Codierung $c_2$	0	10	110	111

Tabelle 2.3: Zwei Beispiele für Codierungen



Abbildung 2.7: Codebäume für die Codierungen  $c_1$  und  $c_2$



**Abbildung 2.7:** Codebäume für die Codierungen  $c_1$  und  $c_2$

- Grundidee Huffmans: bottom-up, Paare mit kleinsten Auftretswahrscheinlichkeiten als Startpunkt

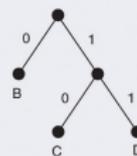
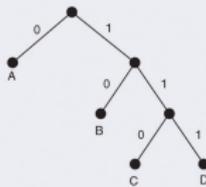
Erster Durchlauf					
Zeichen	A	B	C	D	
Wahrscheinlichkeit	0.5	0.25	0.125	0.125	
Zweiter Durchlauf					
Zeichen/Baum	A	B			
Wahrscheinlichkeit	0.5	0.25	0.25		
Dritter Durchlauf					
Zeichen/Baum	A				
Wahrscheinlichkeit	0.5	0.5			
Vierter Durchlauf					
Zeichen/Baum					
Wahrscheinlichkeit	1.0				

Tabelle 2.7: Ablauf des Huffman-Algorithmus

# Huffman: Beispiel

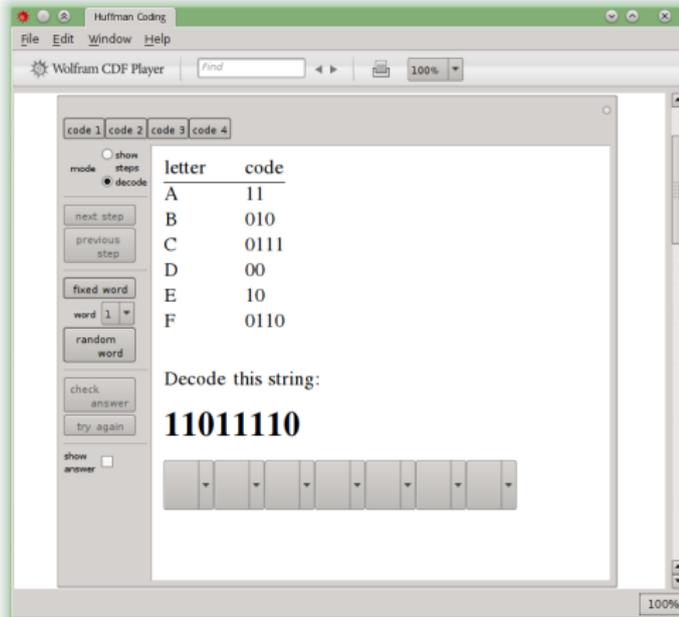
Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren



code 1 code 2 code 3 code 4

show steps  
 decode

letter	code
A	11
B	010
C	0111
D	00
E	10
F	0110

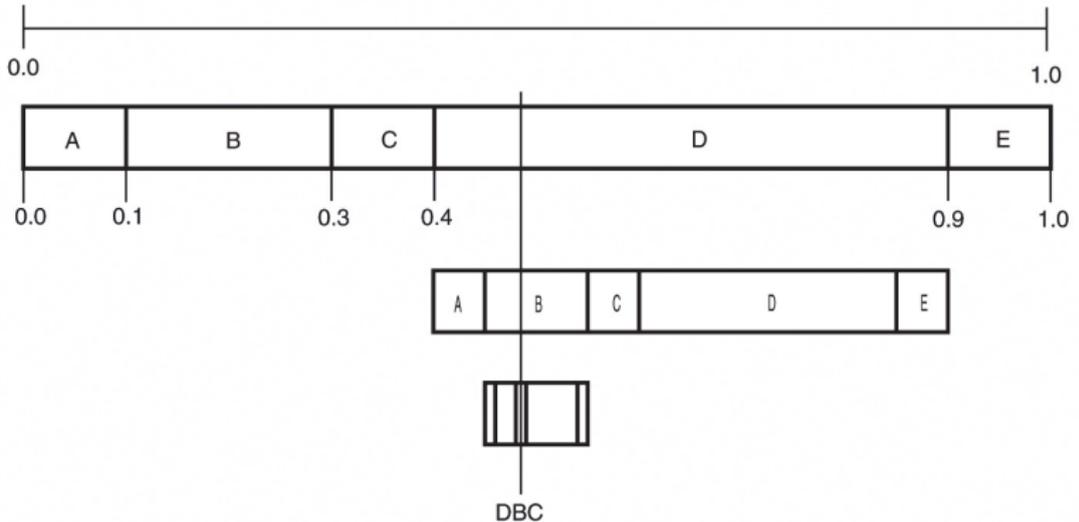
Decode this string:  
**11011110**

▼ ▼ ▼ ▼ ▼ ▼ ▼

100%

[demonstrations.wolfram.com/HuffmanCoding/](https://demonstrations.wolfram.com/HuffmanCoding/)

- Huffman: Ideal wenn Wahrscheinlichkeiten Kehrwerte von Zweierpotenzen
- Entwicklung von Verfahren, die auf beliebigen Wahrscheinlichkeitsverteilungen gute Ergebnisse liefern
- Grundidee Arithmetische Codierung: Ordne einer Nachricht ein Teilintervall aus den reellen Zahlen zwischen 0 und 1 zu
- Länge des Intervalls eine Funktion der Wahrscheinlichkeit des Auftauchens
- bei fortschreitender Codierung werden fortlaufend kleiner Intervalle gebildet
- Wahrscheinlichkeiten der Nachrichtenquelle ergeben erste Intervalleinteilung



**Abbildung 2.8:** Codierung einer Nachricht, die mit „DBC“ beginnt, mit Arithmetischer Codierung

# Andere Verfahren

- Beispiel für zeichenorientierte, universelle verlustfreie Kompression
- Unabhängig von informationstheoretischen Überlegungen
- Grundidee: Speicherung der Wiederholungen eines Zeichens
- AABBBBBBEEDDDDDDDDDDDDB →  
 $\langle A, 2 \rangle \langle B, 5 \rangle \langle E, 2 \rangle \langle D, 11 \rangle \langle B, 1 \rangle$
- Verschiedene Möglichkeiten der Speicherung
  - Dedizierte Trennzeichen
  - Ausnutzen Byte-Struktur (1. Byte Zeichen, 2. Byte Anzahl)

- Lempel-Ziv-Welch-Codierung
- Adaptives Verfahren
  - Starte mit initialem Model
  - Lese die Nachricht
  - Codiere die Nachricht und aktualisiere das Model
- Hier: Wörterbuch für in einer Nachricht vorkommende Teilworte
  - Wörterbuch ist durchnummeriert
  - Nummer übertragen
- Algorithmus speichert nur Teilzeichenfolgen im Wörterbuch, die notwendig sind
- gleichzeitig passiert die Spaltung der Nachricht

- 1 Initialisiere das Wörterbuch mit allen Zeichenketten der Länge 1

# LZW: Ablauf

- 1 Initialisiere das Wörterbuch mit allen Zeichenketten der Länge 1
- 2 Finde den längsten String W im Wörterbuch der derzeit in der Eingabe auftaucht

# LZW: Ablauf

- 1 Initialisiere das Wörterbuch mit allen Zeichenketten der Länge 1
- 2 Finde den längsten String  $W$  im Wörterbuch der derzeit in der Eingabe auftaucht
- 3 Gebe den Wörterbuch-Index für  $W$  aus und entferne  $W$  aus der Eingabe

# LZW: Ablauf

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

- 1 Initialisiere das Wörterbuch mit allen Zeichenketten der Länge 1
- 2 Finde den längsten String  $W$  im Wörterbuch der derzeit in der Eingabe auftaucht
- 3 Gebe den Wörterbuch-Index für  $W$  aus und entferne  $W$  aus der Eingabe
- 4 Füge  $W$  plus dem nächsten Zeichen in der Eingabe dem Wörterbuch hinzu

# LZW: Ablauf

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

- 1 Initialisiere das Wörterbuch mit allen Zeichenketten der Länge 1
- 2 Finde den längsten String  $W$  im Wörterbuch der derzeit in der Eingabe auftaucht
- 3 Gebe den Wörterbuch-Index für  $W$  aus und entferne  $W$  aus der Eingabe
- 4 Füge  $W$  plus dem nächsten Zeichen in der Eingabe dem Wörterbuch hinzu
- 5 Gehe zu Schritt 2

# LZW: Ablauf

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

- 1 Initialisiere das Wörterbuch mit allen Zeichenketten der Länge 1
  - 2 Finde den längsten String  $W$  im Wörterbuch der derzeit in der Eingabe auftaucht
  - 3 Gebe den Wörterbuch-Index für  $W$  aus und entferne  $W$  aus der Eingabe
  - 4 Füge  $W$  plus dem nächsten Zeichen in der Eingabe dem Wörterbuch hinzu
  - 5 Gehe zu Schritt 2
- In der Regel maximale Länge des Wörterbuchs festgelegt
  - Codes länger als zu codierende Zeichen (Länge in Bits), Kompression durch Wiederholung von Substrings

# LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_,

**in**

**b**

**mem**

b

**?**

y

**dict**

none

**out**

nothing

# LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_,

**in**

b

ba

**mem**

b

ba

**?**

y

n

**dict**

none

ba/5

**out**

nothing

1

# LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0

# LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
ban <b>a</b>	na	n	na/7	1,0, <b>3</b>

# LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:na,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
bana	na	n	na/7	1,0,3
ban <b>an</b>	<b>an</b>	y	none	no change

# LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:na,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
bana	na	n	na/7	1,0,3
banan	an	y	none	no change
banana	ana	n	ana/8	1,0,3,6

# LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:na, 8:ana,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
bana	na	n	na/7	1,0,3
banan	an	y	none	no change
banana	ana	n	ana/8	1,0,3,6
banana_	a_	n	a_/9	1,0,3,6,0

# LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:na, 8:ana, 9:a\_,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
bana	na	n	na/7	1,0,3
banan	an	y	none	no change
banana	ana	n	ana/8	1,0,3,6
banana_	a_	n	a_/9	1,0,3,6,0
banana_ <b>b</b>	_b	n	_b/10	1,0,3,6,0, <b>4</b>

## LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:na, 8:ana, 9:a\_, 10:\_b,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
bana	na	n	na/7	1,0,3
banan	an	y	none	no change
banana	ana	n	ana/8	1,0,3,6
banana_	a_	n	a_/9	1,0,3,6,0
banana_b	_b	n	_b/10	1,0,3,6,0,4
banana_ba	ba	y	none	no change

## LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:na, 8:ana, 9:a\_, 10:\_b,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
bana	na	n	na/7	1,0,3
banan	an	y	none	no change
banana	ana	n	ana/8	1,0,3,6
banana_	a_	n	a_/9	1,0,3,6,0
banana_b	_b	n	_b/10	1,0,3,6,0,4
banana_ba	ba	y	none	no change
banana_ban	ban	n	ban/11	1,0,3,6,0,4,5

# LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:na, 8:ana, 9:a\_, 10:\_b, 11:ban,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
bana	na	n	na/7	1,0,3
banan	an	y	none	no change
banana	ana	n	ana/8	1,0,3,6
banana_	a_	n	a_/9	1,0,3,6,0
banana_b	_b	n	_b/10	1,0,3,6,0,4
banana_ba	ba	y	none	no change
banana_ban	ban	n	ban/11	1,0,3,6,0,4,5
banana_ban <b>d</b>	<b>nd</b>	n	nd/12	1,0,3,6,0,4,5, <b>3</b>

## LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:na, 8:ana, 9:a\_, 10:\_b, 11:ban, 12:nd,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
bana	na	n	na/7	1,0,3
banan	an	y	none	no change
banana	ana	n	ana/8	1,0,3,6
banana_	a_	n	a_/9	1,0,3,6,0
banana_b	_b	n	_b/10	1,0,3,6,0,4
banana_ba	ba	y	none	no change
banana_ban	ban	n	ban/11	1,0,3,6,0,4,5
banana_band	nd	n	nd/12	1,0,3,6,0,4,5,3
banana_banda	da	n	da/13	1,0,3,6,0,4,5,3,2

## LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:na, 8:ana, 9:a\_, 10:\_b, 11:ban, 12:nd, 13:da

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
bana	na	n	na/7	1,0,3
banan	an	y	none	no change
banana	ana	n	ana/8	1,0,3,6
banana_	a_	n	a_/9	1,0,3,6,0
banana_b	_b	n	_b/10	1,0,3,6,0,4
banana_ba	ba	y	none	no change
banana_ban	ban	n	ban/11	1,0,3,6,0,4,5
banana_band	nd	n	nd/12	1,0,3,6,0,4,5,3
banana_banda	da	n	da/13	1,0,3,6,0,4,5,3,2
banana_banda <b>n</b>	<b>a</b> n	y	none	no change

# LZW: Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:na, 8:ana, 9:a\_, 10:\_b, 11:ban, 12:nd, 13:da

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
b	b	y	none	nothing
ba	ba	n	ba/5	1
ban	an	n	an/6	1,0
banana	na	n	na/7	1,0,3
banan	an	y	none	no change
banana	ana	n	ana/8	1,0,3,6
banana_	a_	n	a_/9	1,0,3,6,0
banana_b	_b	n	_b/10	1,0,3,6,0,4
banana_ba	ba	y	none	no change
banana_ban	ban	n	ban/11	1,0,3,6,0,4,5
banana_band	nd	n	nd/12	1,0,3,6,0,4,5,3
banana_banda	da	n	da/13	1,0,3,6,0,4,5,3,2
banana_bandan	an	y	none	no change
banana_bandana	ana	y	none	1,0,3,6,0,4,5,3,2,8

# LZW: Beispiel II

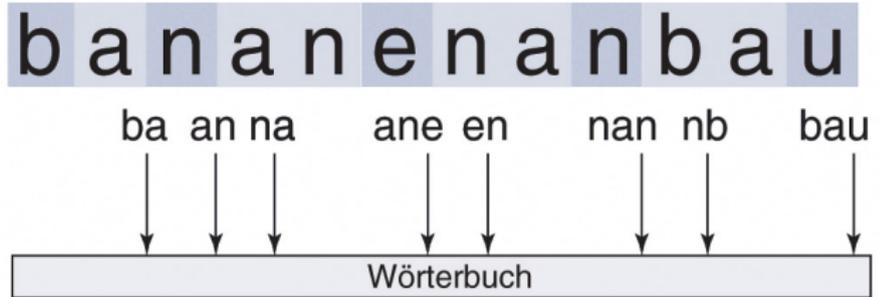
Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren



**Abbildung 2.9:** Ablauf des LZW-Algorithmus an einem Beispiel

# Übung 5.2: LZW Kompression

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

- Initiales Wörterbuch:
  - 0:a, 1:b
- Kodiere:
  - abababab

# LZW: Dekompression

- Nur das Ausgangs-Wörterbuch muß bekannt sein
  - Texte: ASCII?
- Der Rest des Wörterbuches wird bei der Dekompression mit erstellt

# LZW: Dekompression

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

- Nur das Ausgangs-Wörterbuch muß bekannt sein
    - Texte: ASCII?
  - Der Rest des Wörterbuches wird bei der Dekompression mit erstellt
- 
- 1 Lese das nächste Zeichen

# LZW: Dekompression

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

- Nur das Ausgangs-Wörterbuch muß bekannt sein
    - Texte: ASCII?
  - Der Rest des Wörterbuches wird bei der Dekompression mit erstellt
- 
- 1 Lese das nächste Zeichen
  - 2 Sehe im Wörterbuch nach und gebe das Zeichen aus

# LZW: Dekompression

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

- Nur das Ausgangs-Wörterbuch muß bekannt sein
    - Texte: ASCII?
  - Der Rest des Wörterbuches wird bei der Dekompression mit erstellt
- 
- 1 Lese das nächste Zeichen
  - 2 Sehe im Wörterbuch nach und gebe das Zeichen aus
  - 3 Erster Buchstabe des gelesenen Teilwortes wird mit gemerktem Teilwort zu einem neuen Eintrag

- Nur das Ausgangs-Wörterbuch muß bekannt sein
    - Texte: ASCII?
  - Der Rest des Wörterbuches wird bei der Dekompression mit erstellt
- 
- 1 Lese das nächste Zeichen
  - 2 Sehe im Wörterbuch nach und gebe das Zeichen aus
  - 3 Erster Buchstabe des gelesenen Teilwortes wird mit gemerktem Teilwort zu einem neuen Eintrag
  - 4 Dekodiertes Teilwort wird gemerkt

# LZW: Dekompression

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

- Nur das Ausgangs-Wörterbuch muß bekannt sein
    - Texte: ASCII?
  - Der Rest des Wörterbuches wird bei der Dekompression mit erstellt
- 
- 1 Lese das nächste Zeichen
  - 2 Sehe im Wörterbuch nach und gebe das Zeichen aus
  - 3 Erster Buchstabe des gelesenen Teilwortes wird mit gemerktem Teilwort zu einem neuen Eintrag
  - 4 Dekodiertes Teilwort wird gemerkt
  - 5 Gehe zu Schritt 1

# LZW: Dekompression Beispiel I

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

0:a, 1:b, 2:d, 3:n, 4:\_,

**in**

1

**mem ?**

none 1=b

**dict**

none

**out**

b

# LZW: Dekompression Beispiel I

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

0:a, 1:b, 2:d, 3:n, 4:\_,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
1	none	1=b	none	b
1,0	b	0=a	ba/5	ba

# LZW: Dekompression Beispiel I

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
1	none	1=b	none	b
1,0	b	0=a	ba/5	ba
10,3	a	3=n	an/6	ban

# LZW: Dekompression Beispiel I

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
1	none	1=b	none	b
1,0	b	0=a	ba/5	ba
10,3	a	3=n	an/6	ban
103,2	n	2=d	nd/7	band

# LZW: Dekompression Beispiel I

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:nd,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
1	none	1=b	none	b
1,0	b	0=a	ba/5	ba
10,3	a	3=n	an/6	ban
103,2	n	2=d	nd/7	band
1032,6	d	6=an	da/8	band <b>an</b>

# LZW: Dekompression Beispiel I

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:nd, 8:da,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
1	none	1=b	none	b
1,0	b	0=a	ba/5	ba
10,3	a	3=n	an/6	ban
103,2	n	2=d	nd/7	band
1032,6	d	6=an	da/8	bandan
10326,0	an	0=a	ana/9	bandana

# LZW: Dekompression Beispiel I

 Kanäle und  
Medien

Digitalisierung

Kompression

 Stochastische  
Codierung

Andere Verfahren

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:nd, 8:da, 9:ana,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
1	none	1=b	none	b
1,0	b	0=a	ba/5	ba
10,3	a	3=n	an/6	ban
103,2	n	2=d	nd/7	band
1032,6	d	6=an	da/8	bandan
10326,0	an	0=a	ana/9	bandana
103260,4	a	4=_	a_/10	bandana_

# LZW: Dekompression Beispiel I

Kanäle und  
Medien

Digitalisierung

Kompression

Stochastische  
Codierung

Andere Verfahren

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:nd, 8:da, 9:ana, 10:a\_,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
1	none	1=b	none	b
1,0	b	0=a	ba/5	ba
10,3	a	3=n	an/6	ban
103,2	n	2=d	nd/7	band
1032,6	d	6=an	da/8	bandan
10326,0	an	0=a	ana/9	bandana
103260,4	a	4=_	a_/10	bandana_
1032604, <b>5</b>	_	5=ba	_b/11	bandana_ <b>ba</b>

# LZW: Dekompression Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:nd, 8:da, 9:ana, 10:a\_, 11:\_b,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
1	none	1=b	none	b
1,0	b	0=a	ba/5	ba
10,3	a	3=n	an/6	ban
103,2	n	2=d	nd/7	band
1032,6	d	6=an	da/8	bandan
10326,0	an	0=a	ana/9	bandana
103260,4	a	4=_	a_/10	bandana_
1032604,5	_	5=ba	_b/11	bandana_ba
10326045,3	ba	3=n	ban/12	bandana_ban

# LZW: Dekompression Beispiel I

0:a, 1:b, 2:d, 3:n, 4:\_, 5:ba, 6:an, 7:nd, 8:da, 9:ana, 10:a\_, 11:\_b,  
12:ban

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
1	none	1=b	none	b
1,0	b	0=a	ba/5	ba
10,3	a	3=n	an/6	ban
103,2	n	2=d	nd/7	band
1032,6	d	6=an	da/8	bandan
10326,0	an	0=a	ana/9	bandana
103260,4	a	4=_	a_/10	bandana_
1032604,5	_	5=ba	_b/11	bandana_ba
10326045,3	ba	3=n	ban/12	bandana_ban
103260453,9	n	9=ana	na/13	bandana_banana

# LZW: Dekompression Beispiel II

0:a, 1:b,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
0	none	0=a	none	a

## LZW: Dekompression Beispiel II

0:a, 1:b,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
0	none	0=a	none	a
0,1	a	1=b	ab/2	ab

## LZW: Dekompression Beispiel II

0:a, 1:b, 2:ab,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
0	none	0=a	none	a
0,1	a	1=b	ab/2	ab
01,2	b	2=ab	ba/3	ab <b>ab</b>

## LZW: Dekompression Beispiel II

0:a, 1:b, 2:ab, 3:ba,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
0	none	0=a	none	a
0,1	a	1=b	ab/2	ab
01,2	b	2=ab	ba/3	abab
012,4	ab	4=???	???	abab???

## LZW: Dekompression Beispiel II

0:a, 1:b, 2:ab, 3:ba,

<b>in</b>	<b>mem</b>	<b>?</b>	<b>dict</b>	<b>out</b>
0	none	0=a	none	a
0,1	a	1=b	ab/2	ab
01,2	b	2=ab	ba/3	abab
012, <b>4</b>	ab	4=???	???	abab???

- Die Situation, daß ein Eintrag noch nicht im Wörterbuch vorhanden ist, kann nur in einem Fall auftreten:  
Eingabestring der Form abababa (Wiederholung)

## LZW: Dekompression Beispiel II

0:a, 1:b, 2:ab, 3:ba, 4:aba,

in	mem	?	dict	out
0	none	0=a	none	a
0,1	a	1=b	ab/2	ab
01,2	b	2=ab	ba/3	abab
012, <b>4</b>	ab	4=???	???	abab???

- Die Situation, daß ein Eintrag noch nicht im Wörterbuch vorhanden ist, kann nur in einem Fall auftreten:  
Eingabestring der Form **abababa** (Wiederholung)
- Das zuletzt eingetragene Teilwort (aba) wurde gleich nach Erstellung des Eintrags im Wörterbuch benutzt

## LZW: Dekompression Beispiel II

0:a, 1:b, 2:ab, 3:ba, 4:aba,

in	mem	?	dict	out
0	none	0=a	none	a
0,1	a	1=b	ab/2	ab
01,2	b	2=ab	ba/3	abab

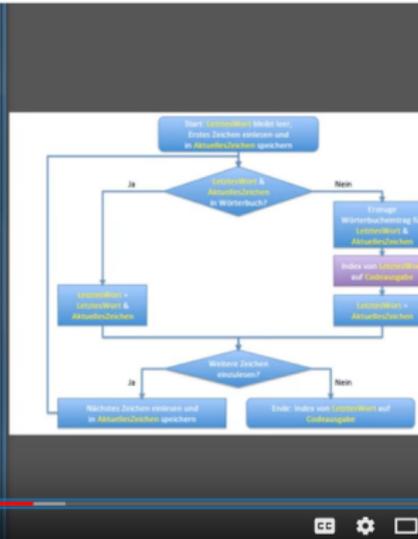
012,4    ab        4=aba    aba/4    abababa

- Die Situation, daß ein Eintrag noch nicht im Wörterbuch vorhanden ist, kann nur in einem Fall auftreten:  
Eingabestring der Form **abababa** (Wiederholung)
- Das zuletzt eingetragene Teilwort (aba) wurde gleich nach Erstellung des Eintrags im Wörterbuch benutzt
- Wir nehmen unseren gemerkten Teilstring (ab) und fügen das erste Zeichen an das Ende an (aba), geben den Teilstring aus und fügen ihn im Wörterbuch hinzu

# Video 5.2: LZW-Kodierung

**Zeichenkette: rabarbarbarbara**

Letztes Wort	Aktuelles Zeichen	Eintrag	Codeausgabe
	r		
r	a	ra = 256	114
a	b	ab = 257	97
b	a	ba = 258	98
a	r	ar = 259	97



ASCI 5:00 / 9:21 b = 98 r = 114

ngocngo tran: LZW Kodierung (9:21)

# Video 5.3: LZW-Dekodierung

Codewort: 114-97-98-97-114-258-260-259-261-97

Letztes Wort	Aktuelles Zeichen	Merke	Codeausgabe
	114		r
r	97	a	a
a	98	b	

Wörterbuch (ASCII)

a = 97	b = 98	r = 114	ra = 256
--------	--------	---------	----------



```

    graph TD
      Start([Start: letztesWort bleibt leer,  
Erste Eingabezahl einlesen und  
in AktuellesZeichen speichern]) --> W1[Wörterbuchbeitrag von AktuellesZeichen  
auf Codeausgabe]
      W1 --> L1[letztesWort = Wörterbuchbeitrag von  
AktuellesZeichen]
      L1 --> N1[Nächste Eingabezahl einlesen und  
in AktuellesZeichen speichern]
      N1 --> W2[Wörterbuchbeitrag von AktuellesZeichen  
auf Merke]
      W2 --> M1[Merke auf Codeausgabe]
      M1 --> E1[Erzeuge neuen Wörterbuchbeitrag für  
letztesWort & 1. Zeichen von Merke]
      E1 --> L2[letztesWort = Merke]
      L2 --> D1{Weitere  
Zahlen  
eingelesen?}
      D1 -- Ja --> N1
      D1 -- Nein --> End([Ende])
  
```

ngocngo tran: LZW Dekodierung (10:57)

# Kanäle, Codecs und Medien

Jörg Cassens

Institut für Mathematik und Angewandte Informatik

Medieninformatik  
WS 2017/2018



medieninformatik

IMA – Institut für  
Mathematik und  
Angewandte Informatik