# Mathematische Grundlagen der Computerlinguistik

## Mengen und Mengenoperationen

Dozentin: Wiebke Petersen

1. Foliensatz

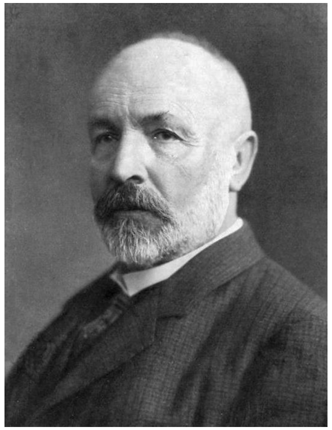


# Mengen

### Georg Cantor (1845-1918)

„Eine Menge ist eine Zusammenfassung von bestimmten wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens (welche die ‚Elemente'genannt werden) zu einem Ganzen." (1895)

- Mengen werden über ihre Elemente bestimmt.
- Elemente von Mengen können selber Mengen sein.
- Mengen können endlich oder unendlich sein.





# Notation und Terminologie

- Variablen für Mengen: $A, B, C, \ldots, M, N, \ldots$
- Variablen für Elemente: $a, b, c, \ldots, x, y, z$
- Ist $m$ ein Element von $M$ so schreibt man $m \in M$.
- Ist $m$ kein Element von $M$ so schreibt man $m \notin M$.
- Zwei Mengen $A$ und $B$ sind genau dann identisch oder gleich, wenn jedes Element von $A$ auch Element von $B$ ist und wenn jedes Element von $B$ auch Element von $A$ ist.
- Es gibt genau eine Menge, die keine Elemente enthält, die leere Menge (Symbol: $\emptyset$, es gilt $\emptyset = \{\ \}$).
- Mengen mit genau einem Element werden Einermengen (*singleton*) genannt.
- $\mathbb{N} = \{1, 2, 3, \ldots\}$ ist die Menge der natürlichen Zahlen
- $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$ ist die Menge der natürlichen Zahlen mit 0
- $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ ist die Menge der ganzen Zahlen
- $\mathbb{Q}$ ist die Menge der rationalen Zahlen (alle ‚Bruchzahlen').
- $\mathbb{R}$ ist die Menge der reellen Zahlen (alle ‚Kommazahlen').



# Bertrand Russell (1872-1970)

### Russels Antinomie (1901)

- Sei $M$ die Menge aller Mengen, die sich nicht selbst als Element enthalten.
- Gilt $M \in M$ oder $M \notin M$?

Ausweg: ‚Theorie der Typen' (Principia Mathematica, Russel & Whitehead 1910-13)

Mengen werden stufenweise aufgebaut und sind immer von einem höheren Typ als ihre Elemente.

## Grellings Paradoxie

Ein Adjektiv heiße

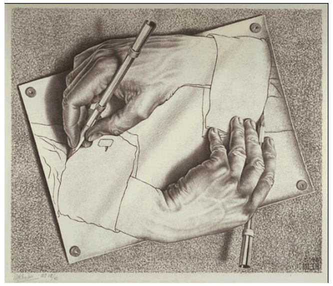
**autologisch,** wenn es sich selbst beschreibt (Bsp.: dreisilbig, kurz, xenonymisch, adjektivisch, verbal, vokalenthaltend, ...)

**heterologisch,** wenn es sich nicht selbst beschreibt (Bsp.: zweisilbig, essbar, grün, ...)

Ist ‚heterologisch' heterologisch?
(nach D.R. Hofstadter: *Gödel, Escher, Bach*)

In diesem Kurs werden Mengen so beschrieben, dass keine Paradoxien auftreten.

Paradoxien der Selbstbezüglichkeit



*zeichnende Hände* von M.C. Escher



## Mengenbeschreibungen

### Explizite Mengendarstellung

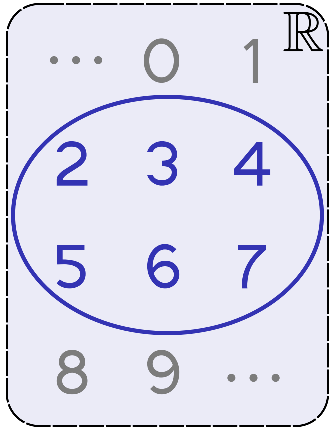
$\{a_1, a_2, \ldots, a_n\}$ ist die Menge, die genau die Elemente $a_1, a_2, \ldots, a_n$ enthält.

Beispiel:
$\{2, 3, 4, 5, 6, 7\}$

### Implizite Mengendarstellung

$\{x|A\}$ ist die Menge, die genau die Objekte $x$ enthält, auf die die Aussage $A$ zutrifft.

Beispiel:
$\{x \in \mathbb{R}|x \in \mathbb{N}$ und $1 < x$ und $x < 8\}$,






## Hinweise zur expliziten Mengendarstellung

- Beschreibung durch Aufzählung oder -listung
- nur für endliche Mengen möglich
- Die Klammern $\{$ und $\}$ heißen Mengenklammern oder geschweifte Klammern.
- Reihenfolge der Elemente spielt keine Rolle: $\{a, b, c\} = \{c, a, b\}$
- Elemente können in der Klammernotation mehrfach auftreten: $\{a, b, c\} = \{a, b, a, b, a, b, c\}$



## Hinweise zur impliziten Mengendarstellung

### Beschreibung mittels charakteristischer Eigenschaft

- { Element $\in$ Grundbereich | Eigenschaft von Element }
- $\{x \in G|E(x)\}$ („Menge aller $x$ in $G$ mit der Eigenschaft $E$")
- Bsp.: $\{x \in \mathbb{N}|x$ ist eine gerade Zahl$\}$
- Wenn der Grundbereich aus dem Kontext bekannt ist oder sich aus der Eigenschaft ergibt, kann er weggelassen werden.
- Bsp.: $\{x|x$ ist eine Primzahl$\}$
- Statt des Symbols '|' verwendet man auch das Symbol ':'. Also $\{x \in \mathbb{N} : x$ ist eine Primzahl$\}$

# Hinweise zur impliziten Mengendarstellung

### Beschreibung mittels rekursiver Definition

Beispiel: Menge der Nachkommen von Georg Cantor

1. Festlegung endlich vieler Startelemente:
   Die Kinder von Cantor sind Nachkommen von Cantor
2. Konstruktionsvorschrift für zusätzliche Elemente:
   Wenn $x$ ein Nachkomme von Cantor ist, dann ist jedes Kind von $x$ ein Nachkomme von Cantor.
3. Einschränkung:
   Nichts sonst ist ein Nachkomme von Cantor.

- Was ist, wenn Cantor keine Kinder hatte?
- Lässt sich so auch die Menge der Nachkommen von Aristoteles definieren? oder die von Merlin?



# Teilmengen

Eine Menge $N$ ist eine Teilmenge der Menge $M$ (in Zeichen: $N \subseteq M$) genau dann, wenn alle Elemente von $N$ auch Elemente von $M$ sind.

- Wenn $x \in N$, dann $x \in M$
- Wenn $y \in M$, dann muss $y \in N$ nicht unbedingt gelten, es kann aber gelten.

Eine Menge $N$ ist eine echte Teilmenge der Menge $M$ (in Zeichen: $N \subset M$) genau dann, wenn $N$ eine Teilmenge von $M$ ist und wenn $M$ und $N$ ungleich sind.

- $N \subseteq M$ und $N \neq M$
- Es gibt ein $y \in M$ mit $y \notin N$.

Wenn $N \subseteq M$, dann ist $M$ eine Übermenge von $N$ (in Zeichen: $M \supseteq N$).

Wenn $M \supseteq N$ und $M \neq N$ dann ist $M$ eine echte Übermenge von $N$ (in Zeichen: $M \supset N$).



# Teilmengen

$x \in M$: $x$ ist ein Element der Menge M

- $2 \in \{1,2,3\}$
- $2 \notin \{1,3,5\}$
- $\{3\} \in \{M | M \text{ ist eine Einermenge}\}$
- $\{3\} \notin \{3\}$

$N \subseteq M$: Die Menge $N$ ist eine Teilmenge der Menge $M$

- $\{2,3\} \subseteq \{1,2,3,4\}$
- $\{2,3\} \subseteq \{2,3\}$
- $\emptyset \subseteq \{1,2,3,4\}$
  (Die leere Menge ist eine Teilmenge jeder Menge!)
- $\{3\} \not\subseteq \{M | M \text{ ist eine Einermenge}\}$

$N \subset M$: Die Menge $N$ ist eine echte Teilmenge der Menge $M$

- $\{1\} \subset \{1,2\}$
- $\{1,2\} \not\subset \{1,2\}$

Vorsicht: Die Element-von- und die Teilmengenrelation müssen streng unterschieden werden!



# Mächtigkeit von Mengen

Zwei Mengen $M$ und $N$ haben dieselbe Mächtigkeit oder heißen gleichmächtig (in Zeichen: $|M| = |N|$), wenn es eine eineindeutige Zuordnung der Elemente von $M$ auf $N$ gibt (d.h., die Zuordnung ordnet jedem Element aus $M$ genau ein Element aus $N$ und jedem Element aus $N$ genau ein Element aus $M$ zu.)

### endliche Mengen

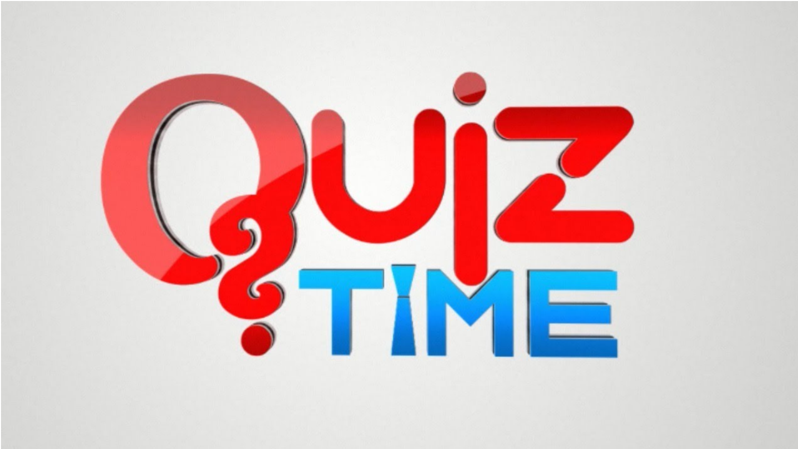
Die Mächtigkeit einer endlichen Menge (in Zeichen: $|M|$) ist die Anzahl ihrer Elemente.

Beispiele:

- $|\emptyset| = 0$
- $|\{1,2\}| = 2$
- $|\{\{1,2\}\}| = 1$

Vorsicht: nicht alle unendlichen Mengen sind gleichmächtig!

## Quiz-Time






## Mengenoperationen (unäre Potenzmengenoperation)

Mengenoperationen sind Abbildungen, die einer oder mehreren Mengen eindeutig eine Menge zuordnen. Einstellige Operationen werden auch unäre und zweistellige auch binäre Operationen genannt.

Die Potenzmengenoperation ist eine unäre Operation, die jeder Menge ihre Potenzmenge zuordnet.

Die Potenzmenge einer Menge $M$ ist die Menge aller möglichen Teilmengen von $M$, also $\mathcal{POT}(M) = \{N | N \subseteq M\}$. Man schreibt auch $2^M$ für die Potenzmenge von $M$.

$$\mathcal{POT}(\{1,2,3\}) = \left\{ \begin{array}{lcccl} \{ & & & & \}, \\ \{ & 1 & & & \}, \\ \{ & & 2 & & \}, \\ \{ & & & 3 & \}, \\ \{ & 1, & 2 & & \}, \\ \{ & 1, & & 3 & \}, \\ \{ & & 2, & 3 & \}, \\ \{ & 1, & 2, & 3 & \}, \end{array} \right\}$$



## Mächtigkeit der Potenzmenge

Für endliche Mengen gilt: ist $M$ eine $n$-elementige Menge, so ist $|\mathcal{POT}(M)| = 2^n$.

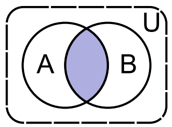
| 1 | 2 | 3 | ... | $n-1$ | $n$ |
|---|---|---|---|---|---|
| 0 | 0 | 0 | ... | 0 | 0 |
| 1 | 0 | 0 | ... | 0 | 0 |
| 0 | 1 | 0 | ... | 0 | 0 |
| 0 | 0 | 1 | ... | 0 | 0 |
| ⋮ | | | | | ⋮ |
| 0 | 0 | 0 | ... | 1 | 0 |
| 0 | 0 | 0 | ... | 0 | 1 |
| 1 | 1 | 0 | ... | 0 | 0 |
| 1 | 0 | 1 | ... | 0 | 0 |
| ⋮ | | | | | ⋮ |
| 1 | 1 | 1 | ... | 1 | 1 |
| $2\times$ | $2\times$ | $2\times$ | ... | $2\times$ | 2 |

$2^n$ Möglichkeiten

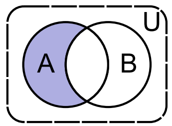


## Mengenoperationen (binäre Operationen)

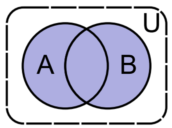
Schnitt: $A \cap B$
„A geschnitten mit B“
$A \cap B = \{x | x \in A \text{ und } x \in B\}$



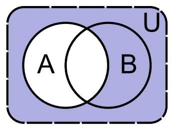

Differenz: $A \setminus B$ (oder $A - B$)
„A ohne B“
$A \setminus B = \{x | x \in A \text{ und } x \notin B\}$




Vereinigung: $A \cup B$
„A vereinigt mit B“
$A \cup B = \{x | x \in A \text{ oder } x \in B\}$




Komplement (in $U$): $C_U(A)$
„Komplement von A in U“
$C_U(A) = U \setminus A$




Wenn $U$ feststeht, schreibt man auch $\overline{A}$

## Mengenoperationen

### Beispiele

$A = \{1,2,3,4\}$, $B = \{3,4,5\}$, $U = \{1,2,3,4,5,6,7\}$,

- $A \cup B = \{1,2,3,4,5\}$, $A \cap B = \{3,4\}$
- $A \setminus B = \{1,2\}$, $\overline{A} = \{5,6,7\}$

### Notation

- Zwei Mengen $A$ und $B$ mit leerem Schnitt heißen disjunkt ($A \cap B = \emptyset$).
- Wenn $A$ eine Menge von Mengen ist, schreiben wir $\bigcup A$ für die Vereinigung aller Elemente von $A$ (Bsp.: $\bigcup\{B, C, D\} = B \cup C \cup D$)
- Wenn $A$ eine Menge von Mengen ist, schreiben wir $\bigcap A$ für den Schnitt aller Elemente von $A$ (Bsp.: $\bigcap\{B, C, D\} = B \cap C \cap D$)
- Häufig werden auch Indizes und Indexmengen zur Notation verwendet. Bsp.: Sei $A_i = \{x \in \mathbb{N}_0 | x \leq i\}$, dann

$$\bigcup_{3 \leq i \leq 5} A_i = \{0,1,2,3,4,5\} \text{ und } \bigcap_{3 \leq i \leq 5} A_i = \{0,1,2,3\}$$



## Eigenschaften der Mengeoperationen (Schnitt und Vereinigung)

### Kommutativgesetze:

$$A \cap B = B \cap A$$
$$A \cup B = B \cup A$$

### Assoziativgesetze:

$$(A \cap B) \cap C = A \cap (B \cap C)$$
$$(A \cup B) \cup C = A \cup (B \cup C)$$

### Distributivgesetze:

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$
$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

### Idempotenzgesetze:

$$A \cap A = A$$
$$A \cup A = A$$

$\emptyset$ ist neutrales Element der Vereinigung: $A \cup \emptyset = \emptyset \cup A = A$

Gibt es auch ein neutrales Element des Schnitts?



## Gesetze der Komplementoperation

### de Morgan:

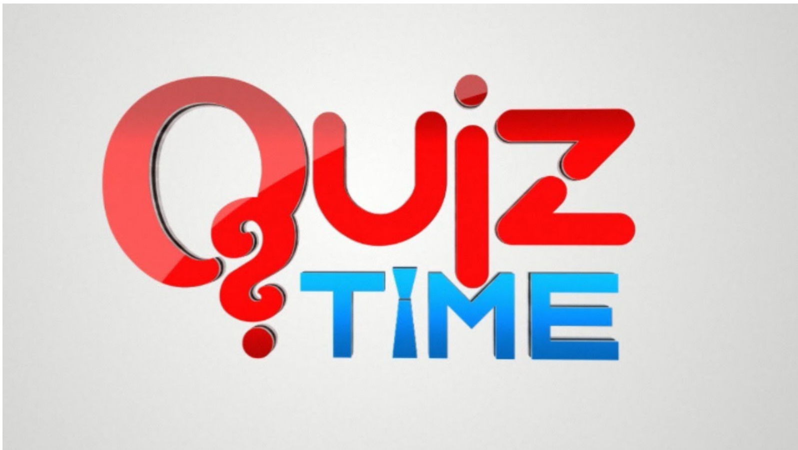
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$
$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

### weitere Gesetze:

$$\overline{\overline{A}} = A$$
$$\overline{A} \cap A = \emptyset$$



## Quiz-Time



# Mathematische Grundlagen der Computerlinguistik

## Relationen und Funktionen

Dozentin: Wiebke Petersen

2. Foliensatz



# $n$-Tupel und Cartesisches Produkt

Mengen sind ungeordnet, häufig werden jedoch geordnete Listen benötigt:

### $n$-Tupel

Ein $n$-Tupel ist eine Liste mit $n \geq 1$ Elementen. Im Gegensatz zu Mengen ist die Reihenfolge festgelegt und jedes Element kann beliebig oft vorkommen.
Beispiel: $\langle 2, 3, 1 \rangle$, $\langle b, e, e, s, i, i, p, l \rangle$
2-Tupel werden auch (geordnete) Paare genannt.

### Cartesisches Produkt

Das Cartesische Produkt (oder Kreuzprodukt) von n Mengen $M_1 \dots M_n$ ist die Menge aller n-Tupel deren i-tes Element aus $M_i$ stammt.
$M_1 \times \dots \times M_n := \{\langle x_1, \dots, x_n \rangle | x_i \in M_i \text{ für } i = 1, \dots, n\}$

Statt $M \times M \times \dots \times M$ schreibt man auch $M^n$, wenn $M$ genau $n$-mal auftritt.

### Beispiel

$M_1 = \{a, b, c\}$, $M_2 = \{a, d\}$
$M_1 \times M_2 = \{\langle a, a \rangle, \langle a, d \rangle, \langle b, a \rangle, \langle b, d \rangle, \langle c, a \rangle, \langle c, d \rangle\}$
$M_1 \times \emptyset = \emptyset$



# Relationen

### Definition

*Eine Teilmenge des Cartesischen Produktes von n Mengen $R \subseteq M_1 \times \dots \times M_n$ heißt n-stellige Relation.*
*Eine Relation $R$ ist also eine Menge von n-Tupeln.*

Hinweis: Relationen werden extensional definiert. Es ist unerheblich, wie die Relation charakterisiert (oder benannt) wird. Wichtig ist allein, welche Objekte zueinander in der Relation stehen.
Für Relationen werden häufig die Buchstaben $R, S, T$ verwendet.

### Beispiele

- Schwester von
- Mutter von
- weibliches Elternteil von
- bilden ein Quartet
- Teilmenge von



# binäre Relationen

- binäre Relationen sind Mengen geordneter Paare
- wenn $a$ in der Relation $R$ zu $b$ steht, dann schreibt man
  - $\langle a, b \rangle \in R$ oder
  - $aRb$ oder
  - $R(a, b)$ oder
  - $Rab$
- Wenn $R \subseteq A \times B$, dann sagt man, dass $R$ eine Relation zwischen $A$ und $B$ ist.
- Wenn $R \subseteq A \times A$, dann sagt man, dass $R$ eine Relation auf $A$ ist.

## inverse und komplementäre Relation

### inverse Relation

Die inverse Relation zu einer binären Relation $R \subseteq A \times B$ ist die Relation
$R^{-1} = \{\langle b, a\rangle \in B \times A | \langle a, b\rangle \in R\}$.

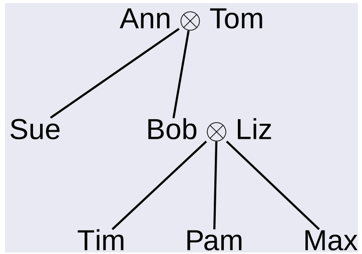
### komplementäre Relation

Die komplementäre Relation zu einer binären Relation $R \subseteq A \times B$ zwischen $A$ und $B$ ist die Relation
$R' = A \times B \setminus R$.



## Beispiel: Verwandtschaftsterme

### Beispielfamilie



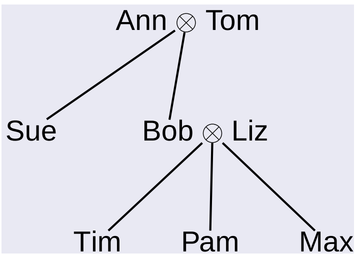

### ‚hat als Sohn'

| | | |
|---|---|---|
| Ann | $R_{\text{son}}$ | Bob |
| Tom | $R_{\text{son}}$ | Bob |
| Bob | $R_{\text{son}}$ | Max |
| Bob | $R_{\text{son}}$ | Tim |
| Liz | $R_{\text{son}}$ | Max |
| Liz | $R_{\text{son}}$ | Tim |



## Beispiel: Verwandtschaftsterme

### Beispielfamilie




### ‚hat als Mutter'

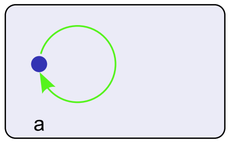
| | | |
|---|---|---|
| Sue | $R_{\text{mother}}$ | Ann |
| Bob | $R_{\text{mother}}$ | Ann |
| Tim | $R_{\text{mother}}$ | Liz |
| Pam | $R_{\text{mother}}$ | Liz |
| Max | $R_{\text{mother}}$ | Liz |



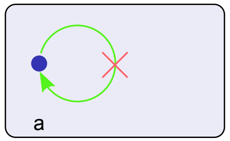
## Eigenschaften binärer Relationen

Sei $R \subseteq A \times A$ eine binäre Relation auf $A$.

$R$ ist reflexiv g.d.w. für alle $a \in A$ gilt, dass $aRa$.




$R$ ist irreflexiv g.d.w. für kein $a \in A$ gilt, dass $aRa$

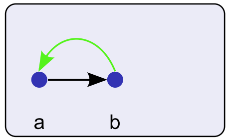



- Die Relation ‚hat am selben Tag Geburtstag' auf der Menge der Menschen ist reflexiv.
- Die Relation ‚ist Mutter von' auf der Menge der Menschen ist irreflexiv.
- Die Relation ‚kann die Quersumme des Geburtstags von berechnen' auf der Menge der Menschen ist weder reflexiv noch irreflexiv.
- Welche Bedingungen erfüllen die Beispielrelationen an der Tafel?

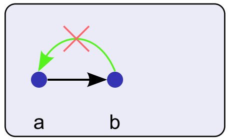
## Eigenschaften binärer Relationen

Sei $R \subseteq A \times A$ eine binäre Relation auf $A$.

$R$ ist symmetrisch g.d.w. für alle $a, b \in A$ mit $aRb$ auch $bRa$ gilt.




$R$ ist asymmetrisch g.d.w. für $a, b \in A$ niemals sowohl $aRb$ als auch $bRa$ gilt.




$R$ ist antisymmetrisch g.d.w. für alle $a, b \in A$ aus $aRb$ und $bRa$ folgt, dass $a = b$.

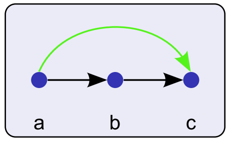
- Die Relation ‚ist verheiratet mit' ist symmetrisch.
- Die Relation ‚ist größer als' ist asymmetrisch.
- Die Relation ‚ist Teilmenge von ' ist antisymmetrisch.
- Welche Bedingungen erfüllen die Beispielrelationen an der Tafel?



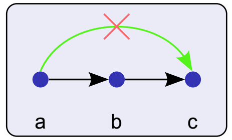
## Eigenschaften binärer Relationen

Sei $R \subseteq A \times A$ eine binäre Relation auf $A$.

$R$ ist transitiv g.d.w. für alle $a, b, c \in A$ aus $aRb$ und $bRc$ immer $aRc$ folgt.




$R$ ist intransitiv g.d.w. für alle $a, b, c \in A$ mit $aRb$ und $bRc$ niemals $aRc$ gilt.




- Die Relation ‚ist Vorfahr von' ist transitiv.
- Die Relation ‚steht genau eine Treppenstufe höher als' ist intransitiv.
- Die Relation ‚kennt' ist weder transitiv noch intransitiv.
- Welche Bedingungen erfüllen die Beispielrelationen an der Tafel?



## Definitions- und Wertebereich einer Relation

Wenn $R \subseteq A \times B$ eine binäre Relation ist, dann heißt

$$\mathrm{dom}(R) = \{a \in A \mid \text{ es gibt ein } b \in B \text{ mit } (a, b) \in R\}$$

der Definitionsbereich (domain) von $R$.
Die Menge

$$\mathrm{rng}(R) = \{b \in B \mid \text{ es gibt ein } a \in A \text{ mit } (a, b) \in R\}$$

heißt der Wertebereich (range) von $R$.

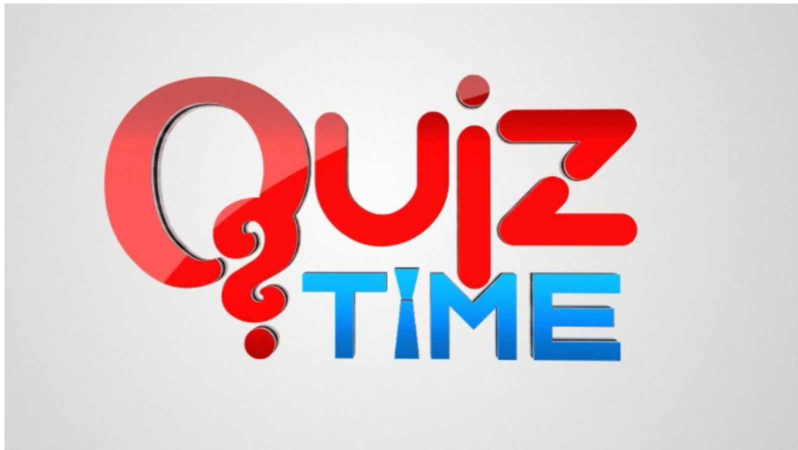
Beispiel:

$$A = \{a, b, c, d\},\ B = \{1, 2, 3, 4, 5\},\ R = \{(b, 1), (b, 2), (c, 3)\}$$

$$\mathrm{dom}(R) = \{b, c\},\ \mathrm{rng}(R) = \{1, 2, 3\}$$



## Quiz-Time



# Äquivalenzrelation

### Äquivalenzrelation

Eine Relation $R \subseteq A \times A$ ist eine Äquivalenzrelation auf $A$, g.d.w. $R$ reflexiv, symmetrisch und transitiv ist.

Wenn $R$ eine Äquivalenzrelation ist und $aRb$ gilt, dann sagt man, dass $a$ äquivalent ist zu $b$ bezüglich $R$.

Für Äquivalenzrelationen verwendet man häufig das Symbol $\sim$.

Beispiele:

- Gleichheit
- ist im selben Semester wie
- hat gleich viele Elemente wie
- hat die selbe Farbe wie
- Welche der Beispielrelationen an der Tafel sind Äquivalenzrelationen?



# Äquivalenzrelation

### Äquivalenzklasse

Sei $R$ eine Äquivalenzrelation auf $A$. Die Äquivalenzklasse eines Elements $a \in A$ ist die Menge aller zu $a$ äquivalenten Elemente von $A$, also

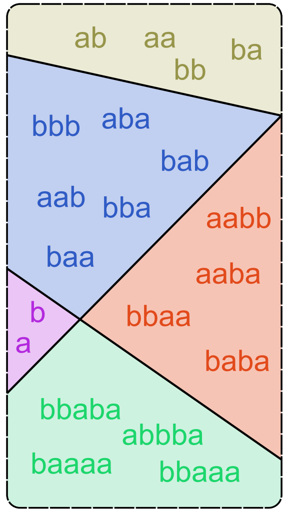
$$[a]_R = \{b \in A | aRb\}.$$

Die Menge

$$A/R = \{[a]_R | a \in A\}$$

aller Äquivalenzklassen von Elementen aus $A$ bezüglich $R$ heißt Quotient von $A$ bezüglich $R$.

Hinweis: Äquivalenzklassen können per Definition nicht leer sein.





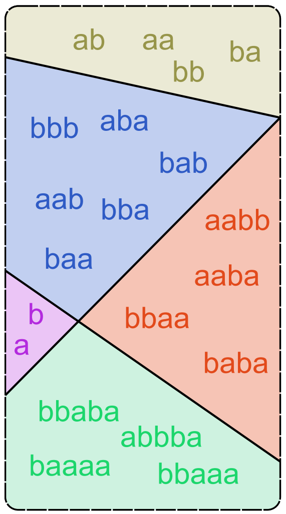

# Äquivalenzrelation

Sei $R$ eine Äquivalenzrelation auf $A$. Dann gilt:

- Zwei Äquivalenzklassen von $R$ sind entweder disjunkt oder identisch: für alle $a, b \in A$ gilt entweder $[a]_R \cap [b]_R = \emptyset$ oder $[a]_R = [b]_R$.
- Die Äquivalenzklassen von $R$ decken ganz $A$ ab: $\bigcup A/R = A$.

Eine Menge $P \subseteq \mathcal{POT}(A)$ ist eine Partition (oder disjunkte Zerlegung) von $A$, g.d.w. $\bigcup P = A$ und für alle $X, Y \in P$ mit $X \neq Y$ gilt $X \cap Y = \emptyset$.

Folglich bildet der Quotient einer Äquivalenzrelation eine Partition der Grundmenge.






# Quiz-Time

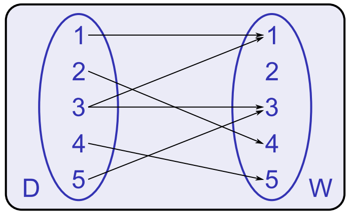
# Funktionen

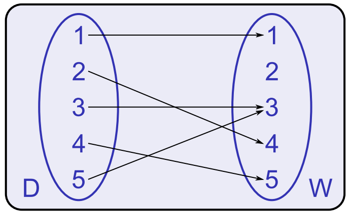
### Definition

*Eine Relation $R \subseteq D \times W$ ist eine Funktion (oder Abbildung), wenn sie jedem Element aus $D$ genau ein Element aus $W$ zuordnet.*
*Funktionen müssen also die Bedigungen der Existenz und Eindeutigkeit erfüllen:*

*Existenz: Für **jedes** $x \in D$ gibt es ein $y \in W$ mit $\langle x, y \rangle \in R$.*

*Eindeutigkeit: Wenn $\langle x, y \rangle \in R$ und $\langle x, z \rangle \in R$, dann $y = z$.*




Eine Relation, für die die Eindeutigkeitsbedingung (aber nicht unbedingt die Existenzbedingung) gilt, heißt partielle Funktion.



# Notation und Terminologie

- Für Funktionen verwendet man häufig die Buchstaben $f, g, h, F, G, H$.
- Wenn $f \subseteq A \times B$ eine Funktion ist, dann sagt man, dass $f$ eine Funktion von $A$ nach $B$ ist, und schreibt $f : A \to B$. $A$ wird dann der Definitionsbereich und $B$ der Wertebereich von $f$ genannt.
- Wenn $\langle a, b \rangle \in f$, dann sagt man, dass die Funktion $f$ dem Element $a$ den Wert $b$ zuweist, und schreibt $f(a) = b$ oder $f : a \mapsto b$.
- Elemente des Definitionsbereiches heißen Argumente und Elemente des Wertebereiches heißen Werte einer Funktion.
- Wenn $C \subset A$ und $f : A \to B$, dann bezeichnet $f|_C : C \to B$ die Einschränkung der Funktion $f$ auf $C$. Für alle $c \in C$ gilt $f|_C(c) = f(c)$.
- Im Kontext von partiellen Funktionen werden Funktionen, die die Existenzbedingung erfüllen, häufig totale Funktionen genannt.



# Beispiele

Sei $A = \{a, b, c, d\}$ $B = \{1, 2, 3, 4, 5\}$

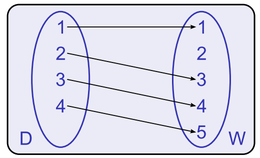
- Die Relation $R \subseteq A \times B$ mit $R = \{(b, 1), (b, 2), (c, 3)\}$ ist keine partielle Funktion.
- Die Relation $R \subseteq A \times B$ mit $R = \{(b, 1), (c, 3), (d, 1)\}$ ist eine partielle aber keine totale Funktion.
- Die Relation $R \subseteq A \times B$ mit $R = \{(a, 2), (b, 1), (c, 3), (d, 1)\}$ ist eine totale und folglich auch eine partielle Funktion.



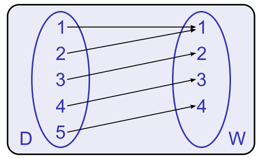
# Funktionseigenschaften

Sei $f : D \to W$ eine Funktion.

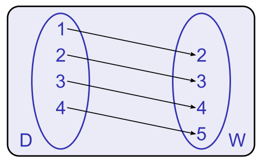
$f$ ist injektiv (Engl.: one-to-one), wenn keine zwei verschiedenen Elemente des Definitionsbereiches denselben Wert zugewiesen bekommen. Wenn also für alle $x, y \in D$ gilt:
$f(x) = f(y)$ g.d.w. $x = y$.




$f$ ist surjektiv (Engl.: onto), wenn jedes Element von $W$ mindestens einem Element von $D$ als Wert zugewiesen wird. Wenn es also für jedes $y \in W$ ein $x \in D$ gibt, für das $f(x) = y$ gilt.

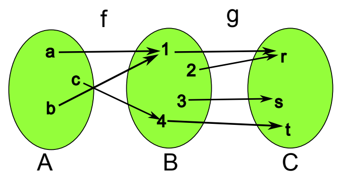


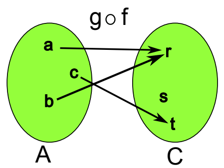

$f$ ist bijektiv, wenn $f$ *injektiv* **und** *surjektiv* ist.
Merke: Eine Funktion $f$ ist bijektiv, g.d.w. $f^{-1}$ eine Funktion ist.



# Komposition von Funktionen

Seien $f : A \to B$ und $g : B \to C$ zwei Funktionen. Die Funktion $g \circ f : A \to C$ mit $g \circ f = \{(x, z) \in A \times C \mid$ es gibt ein $y \in B$ mit $(x, y) \in f$ und $(y, z) \in g\}$ ist die Komposition (oder Verkettung) von $f$ und $g$.
Es gilt $(g \circ f)(x) = g(f(x))$. Die Funktion $g \circ f$ weist einem Element $x \in A$ das Element aus $C$ zu, das man erhält, wenn man zunächst $f$ auf $x$ anwendet und auf das Ergebnis noch $g$ anwendet.

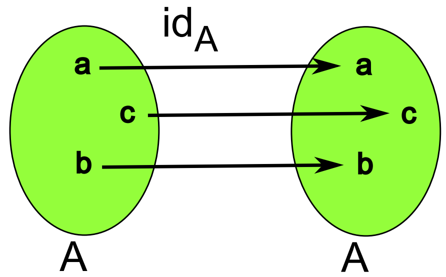





# Identitätsfunktion

Die Funktion $id_A : A \to A$ mit $f = \{(a, a) \in A \times A\}$ (oder $f(a) = a$ für alle $a \in A$) heißt die Identität(sfunktion) auf $A$.






# mehrstellige Funktionen

- Der Definitionsbereich einer Funktion kann selbst eine Relation sein.
- Eine Funktion $A_1 \times A_2 \times \ldots \times A_n \to B$ heißt $n$-stellige Funktion.
- Beispiel: Die Addition der natürlichen Zahlen $+ : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{N}_0$ kann als zweistellige Funktion aufgefasst werden.
- Zweistellige Operationen bilden zweistellige Funktionen (Bsp.: Schnitt, Vereinigung, ...).
- $n$-stellige Funktionen sind $n + 1$-stellige Relationen (Bsp: Mutter)



# Charakteristische Funktion einer Teilmenge

Eine Teilmenge $N \subseteq M$ lässt sich mithilfe ihrer charakteristischen Funktion beschreiben.
Die charakteristische Funktion einer Teilmenge $N \subseteq M$ ist die Funktion $\chi : M \to \{0, 1\}$, für die gilt: $\chi(x) = 1$ genau dann, wenn $x \in N$.
Für die charakteristische Funktion von $N \subseteq M$ schreibt man häufig auch $\chi_N$.
Es gilt:

$$\chi_N : M \to \{0, 1\}; \quad \chi_N(x) = \begin{cases} 1 & \text{wenn} x \in N \\ 0 & \text{sonst} \end{cases}$$

## Mengen von Funktionen

Mit $M^N$ bezeichnet man die Menge aller Funktionen von $N$ nach $M$. Also:

$$M^N = \{f : N \to M \mid f \text{ ist eine Funktion}\}$$



## Charakteristische Funktion und Potenzmenge

Wir haben gesehen, dass man für die Potenzmenge einer Menge $M$ auch $2^M$ schreiben kann. Warum?
In $2^M$ steht 2 für die 2-elementige Menge $\{0, 1\}$.
Die Potenzmenge einer Menge $M$ lässt sich als Menge aller charakteristischen Funktionen ihrer Teilmengen auffassen:

$$\mathcal{POT}(M) = 2^M = \{f : M \to \{0, 1\} \mid f \text{ ist eine Funktion}\}$$

| 1 | 2 | 3 | ... | n |
|---|---|---|---|---|
| 0 | 0 | 0 | ... | 0 |
| 1 | 0 | 0 | ... | 0 |
| 0 | 1 | 0 | ... | 0 |
| ⋮ | | | | ⋮ |
| 0 | 0 | 0 | ... | 1 |
| 1 | 1 | 0 | ... | 0 |
| 1 | 0 | 1 | ... | 0 |
| ⋮ | | | | ⋮ |
| 1 | 1 | 1 | ... | 1 |



## Quiz-Time




# Mathematische Grundlagen der Computerlinguistik

## formale Sprachen

Dozentin: Wiebke Petersen

3. Foliensatz

## Alphabete und Wörter

### Definition

- Alphabet $\Sigma$: endliche Menge von Symbolen / Zeichen.
- Wort: eine endliche Kette/Folge $x_1 \dots x_n$ von Symbolen/Zeichen eines Alphabets (mit $n \geq 0$). Das Wort, das aus null Zeichen besteht heißt leeres Wort und wird mit $\varepsilon$ bezeichnet.
- Die Menge aller Wörter über einem Alphabet $\Sigma$ bezeichnen wir mit $\Sigma^*$.

  $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$ ist die Menge der nichtleeren Wörter.
- Länge eines Wortes $|w|$: Gesamtzahl der Zeichen eines Wortes $w$ ($|abbaca| = 6$, $|\varepsilon| = 0$)



## Leersymbol, leeres Wort und leere Menge

### Vorsicht Verwechslungsgefahr!

Das Leersymbol ␣ ist ein *Zeichen* des Alphabets, also ist ein Wort, das nur aus dem Leersymbol besteht, ein Wort der Länge 1.

Das leere Wort $\varepsilon$ ist ein *Wort* der Länge 0.

Die leere Menge $\emptyset$ ist eine *Menge*.



## Übung: Alphabete und Wörter

Sei $\Sigma = \{a, b, c\}$ ein Alphabet:

- Geben Sie zwei Wörter der Läge 4 über $\Sigma$ an.
- Welche der folgenden Ausdrücke sind Wörter über $\Sigma$ und welche Länge haben sie?: 
  '$aa$', '$caab$', '$da$'
- Was ist der Unterschied zwischen $\Sigma^*$, $\Sigma^+$ und $\Sigma$?
- Wieviele Elemente haben $\Sigma$, $\Sigma^*$ und $\Sigma^+$?



## Operationen auf Wörtern

### Verkettung / Konkatenation

Die Konkatenation / Verkettung zweier Wörter $u = a_1 a_2 \dots a_n$ und $v = b_1 b_2 \dots b_m$ mit $n, m \geq 0$ ist

$$u \circ v = a_1 \dots a_n b_1 \dots b_m$$

Häufig schreiben wir $uv$ statt $u \circ v$.

$$w \circ \varepsilon = \varepsilon \circ w = w \quad \text{Neutrales Element}$$

$$u \circ (v \circ w) = (u \circ v) \circ w \quad \text{Assoziativität}$$

Ist die Konkatenationsoperation kommutativ?

# Symbolpolitik der Mathematik

Vorsicht:

- Obwohl die Symbole für die Komposition von Funktionen und die Konkatenation von Wörtern übereinstimmen, handelt es sich um unterschiedliche Operationen!
- In der Mathematik finden sie häufig mehrdeutige Symbole, deren Bedeutung sich aus dem jeweiligen Kontext ergibt.
- Sie müssen sich also bei dem Symbol $\circ$ immer fragen, ob es zwischen Funktionen oder Wörtern steht (wir werden auch noch eine Operation auf Mengen kennenlernen, die mit demselben Symbol bezeichnet wird).
- Bedenken Sie, dass die Alternative die Verwendung einer unbegrenzten Zahl verschiedener Symbole wäre, da es theoretisch unendlich viele Operationen gibt. Jedes dieser Symbole müsste in Zeichensätzen vorgehalten werden, was unmöglich ist, da Alphabete endlich sein müssen. Stellen Sie sich außerdem vor, ich würde an der Tafel versuchen eine Vielzahl von sehr ähnlichen Symbolen zu verwenden (Beispiel: Kreis mit dickem Punkt in der Mitte, Kreis mit kleinem Punkt, Kreis ohne Punkt, Kreis mit zwei Umrandungen,...), Sie würden das nicht lesen wollen!



# Operationen auf Wörtern

**Exponenten**

- $w^n$: $w$ wird $n$-mal mit sich selbst verkettet.
- $w^0 = \varepsilon$ : $w$ wird '0-mal' mit sich selbst verkettet.

**Umkehrung**

- Die Umkehrung eines Wortes $w$ wird mit $w^R$ bezeichnet. $(abcd)^R = dcba$.
- Ein Wort $w$, für das $w = w^R$ gilt, heißt Palindrom.

(madam, reliefpfeiler, otto, anna, regallager ...)



# Übung: Operationen auf Wörtern

Seien $w = aabc$ und $v = bcc$ Wörter, ermitteln Sie:

- $w \circ v$
- $((w^R \circ v)^R)^2$
- $w \circ (v^R \circ w^3)^0$



# Formale Sprache

**Definition**

Eine formale Sprache $L$ ist eine Menge von Wörtern über einem Alphabet $\Sigma$, also $L \subseteq \Sigma^*$.

Beispiele:

- Sprache $L_{rom}$ der gültigen römischen Zahldarstellungen über dem Alphabet $\Sigma_{rom} = \{\mathbf{I}, \mathbf{V}, \mathbf{X}, \mathbf{L}, \mathbf{C}, \mathbf{D}, \mathbf{M}\}$.
- Sprache $L_{Mors}$ der Buchstaben des lateinischen Alphabets dargestellt im Morsecode. $L_{Mors} = \{\cdot -, - \cdot \cdot\cdot, \dots, - - \cdot\cdot\}$
- Sprache $L_{pal}$ der Palindrome im deutschen Duden $L_{pal} = \{\text{Madam, reliefpfeiler, } \dots\}$
- Leere Menge
- Menge der Wörter der Länge 13 über dem Alphabet $\{a, b, c\}$
- Sprache der syntaktisch wohlgeformten Python-Programme
- Deutsch?

## Operationen auf Sprachen

Seien $L \subseteq \Sigma^*$ und $K \subseteq \Sigma^*$ zwei Sprachen über dem Alphabet $\Sigma$, dann entstehen durch die Verknüpfung mit Mengenoperatoren neue Sprachen über $\Sigma$:

$$K \cup L,\ K \cap L,\ K \setminus L$$

Die Verkettung von Wörtern kann ausgedehnt werden auf die Verkettung von Sprachen:

$$K \circ L := \{v \circ w \in \Sigma^* | v \in K, w \in L\}$$

Beispiel: Sei $K = \{abb, a\}$ und $L = \{bbb, ab\}$

- $K \circ L = \{abbbbb, abbab, abbb, aab\}$ und $L \circ K = \{bbbabb, bbba, ababb, aba\}$
- $K \circ \emptyset = \emptyset$
- $K \circ \{\varepsilon\} = K$
- $K^2 = K \circ K = \{abbabb, abba, aabb, aa\}$



## Potenzen von Sprachen, Iteration, Kleene-Stern

Die $n$-te Potenz einer Sprache $L$ ist die $n$-fache Verkettung von $L$ mit sich selbst:

$$L^n = \underbrace{L \circ L \circ L \ldots \circ L}_{n\text{-mal}}$$

Induktive Definition:

$$L^0 = \{\epsilon\}, \quad L^{n+1} = L^n \circ L$$

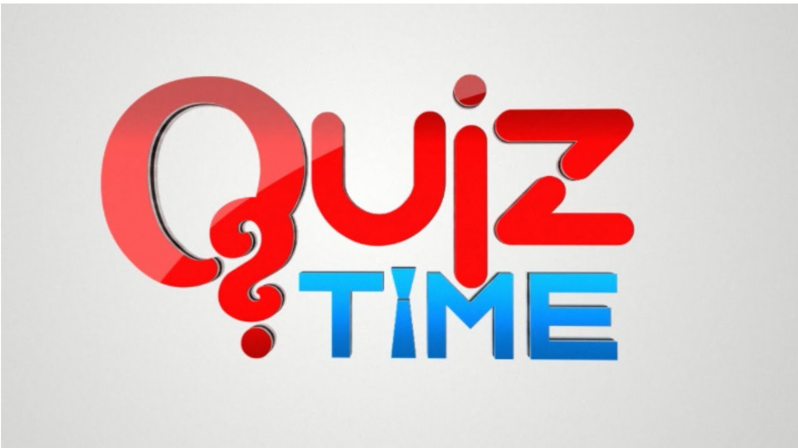
Die Iteration (Kleene-Stern) von $L$ ist

$$L^* := \bigcup_{n \geq 0} L^n$$

Für jede beliebige Sprache $L$ gilt: $\epsilon \in L^*$
Also gilt: $\emptyset^* = \{\epsilon\}$



## Quiz-Time



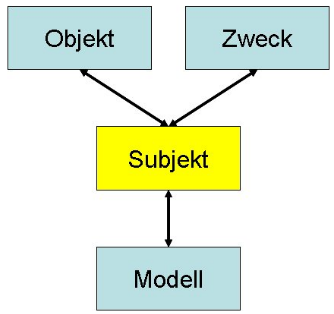



## Modell

- künstlich geschaffen
- materiell oder immateriell
- vereinfachtes Abbild
- zweckgerichtet
- Abstraktion
- Repräsentation
- Modellierungsannahmen

### Modellierung

Ein Subjekt entwirft zu einem Original ein Modell zu einem bestimmten Zweck.



# Modellierung natürlicher Sprachen

## Formale Sprachen

Formale Sprachen sind Mengen von Wörtern (entspricht in natürlichen Sprachen den **Sätzen**), die ihrerseits aus Zeichen/Symbolen (in natürlichen Sprachen **Wörtern**) aufgebaut sind. Was in der Menge ist, ist ein "grammatisch korrektes Wort", alles andere nicht.

Für "strukturierte" formale Sprachen lassen sich endliche Mengen von Regeln/Grammatiken angeben, die diese beschreiben.

## Sprachmodell

Formale Sprachen dienen als Modell für natürliche Sprachen.

Wir gehen davon aus, daß alle natürlichen Sprachen durch endlich viele Regeln beschreibbar sind, da wir sie ansonsten nicht sprechen / verstehen könnten.

Welche Modellannahmen werden hier implizit gemacht?



# Sprachbeschreibung durch Aufzählung aller Wörter

- Peter says that Mary has fallen off the tree.
- Oskar says that Peter says that Mary has fallen off the tree.
- Lisa says that Oskar says that Peter says that Mary has fallen off the tree.
- ...

Scheitert bei unendlichen Sprachen.

Aufzählungen erfassen keine Generalisierungen.



# Sprachbeschreibung durch Angabe einer Grammatik

## Grammatik

- Eine formale Grammatik ist ein generativer Mechanismus zur Erzeugung von Zeichenketten.
- Grammatiken sind endliche Regelsysteme.
- Die Menge aller Ketten, die von einer Grammatik generiert werden, bilden die von der Grammatik beschriebene formale Sprache.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| S | $\rightarrow$ | NP VP | VP | $\rightarrow$ | V | VP | $\rightarrow$ | VP and VP |
| NP | $\rightarrow$ | D N | NP | $\rightarrow$ | NP or NP | D | $\rightarrow$ | the |
| N | $\rightarrow$ | cat | N | $\rightarrow$ | dog | V | $\rightarrow$ | sleeps |
| V | $\rightarrow$ | dreams | | | | | | |

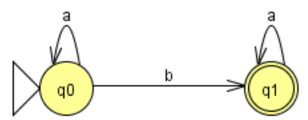
Generiert: the cat sleeps, the dog sleeps, the cat sleeps and dreams,...
aber auch: the cat or the dog sleeps and dreams, ...



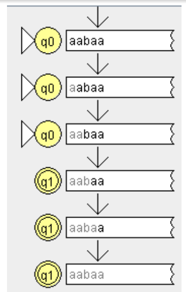
# Sprachbeschreibung durch Automaten

## Automaten

- Ein Automat ist eine abstrakte Maschine, die bestimmte Zeichenketten akzeptiert.
- Die Menge aller Ketten, die von einem Automaten akzeptiert werden, bilden die von dem Automaten beschriebene formale Sprache.




akzeptiert die Sprache $\{a\}^* \circ \{b\} \circ \{a\}^*$



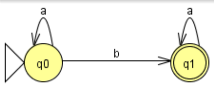
# einfachstes Automatenmodell: endliche Automaten

## Definition

Ein endlicher Automat ist ein 5-Tupel $(Q, \Sigma, \Delta, q_0, F)$ bestehend aus:

1. $Q$: Alphabet der Zustände
2. $\Sigma$: Eingabealphabet ($Q$ und $\Sigma$ müssen disjunkt sein)
3. $\Delta$: Übergangsrelation ($\Delta \subseteq Q \times \Sigma \times Q$)
4. $q_0$: Startzustand ($q_0 \in Q$)
5. $F$: Menge der Endzustände $F \subseteq Q$.

Der Automat heißt deterministisch, wenn die Übergangsrelation $\Delta$ eine (partielle) Funktion ist ($\Delta : Q \times \Sigma \to Q$).

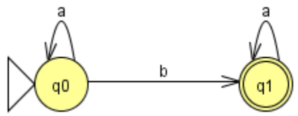


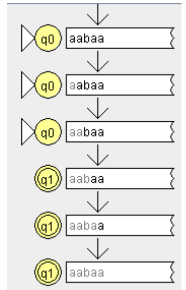



# endliche Automaten: Akzeptanz von Wörtern

Ein endlicher Automat akzeptiert ein Wort $w$, wenn es möglich ist

- beginnend im Startzustand
- das Wort Symbol für Symbol abzuarbeiten, indem man den Zustand gemäß der Übergangsrelation wechselt
- bis das Wort vollständig abgearbeitet ist,
- und wenn man sich am Ende in einem Endzustand befindet.





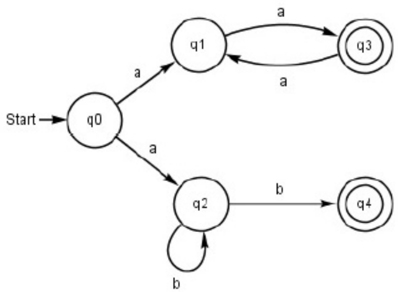

# Beispiel: endlicher Automat

als 5-Tupel:

$(Q, \Sigma, \Delta, q_0, F)$ mit

- $Q = \{q_0, q_1, q_2, q_3, q_4\}$
- $\Sigma = \{a, b\}$
- $\Delta = \{(q_0, a, q_1), (q_0, a, q_2), (q_1, a, q_3), (q_3, a, q_1), (q_2, b, q_2), (q_2, b, q_4)\}$
- $F = \{q_3, q_4\}$

als Übergangsnetz:




Dieser Automat ist nicht deterministisch
(am Übergangsnetz ablesbar an identisch beschrifteten Kanten, die von demselben Knoten ausgehen)



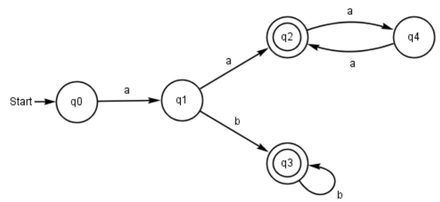
# Beispiel: endlicher Automat

als 5-Tupel:

$(Q, \Sigma, \Delta, q_0, F)$ mit

- $Q = \{q_0, q_1, q_2, q_3, q_4\}$
- $\Sigma = \{a, b\}$
- $\Delta = \{(q_0, a, q_1), (q_1, a, q_2), (q_1, b, q_3), (q_3, b, q_3), (q_2, a, q_4), (q_4, a, q_2)\}$
- $F = \{q_2, q_3\}$

als Übergangsnetz:




Dieser Automat ist deterministisch und akzeptiert dieselbe Sprache wie der Automat der vorangegangenen Folie, nämlich $\{a\} \circ ((\{a\} \circ (\{a\} \circ \{a\})^*) \cup (\{b\} \circ \{b\}^*))$.
Dies ist die Sprache aller Wörter über dem Alphabet $\{a, b\}$, die aus einem $a$ gefolgt von einer beliebigen, nichtleeren Kette von $b$'s oder aus einer nichtleeren Kette von $a$'s gerader Länge bestehen.

# Endliche Automaten: Terminologie

- Zwei Automaten, die dieselbe Sprache akzeptieren, heißen äquivalent (Beispiel: die Automaten der letzten beiden Folien sind äquivalent)
- Satz: Zu jedem nichtdeterministischen endlichen Automaten gibt es einen äquivalenten deterministischen endlichen Automaten.
- Übergangsrelationen werden häufig als Übergangstabellen dargestellt.
Beispiel: $\Delta = \{(q_0, a, q_1), (q_1, a, q_2), (q_1, b, q_3), (q_3, b, q_3), (q_2, a, q_4), (q_4, a, q_2)\}$

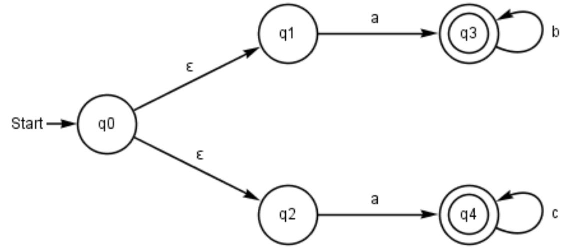
| | a | b |
|---|---|---|
| $q_0$ | $q_1$ | |
| $q_1$ | $q_2$ | $q_3$ |
| $q_2$ | $q_4$ | |
| $q_3$ | | $q_3$ |
| $q_4$ | $q_2$ | |

- Ist die Übergangsrelation eines endlichen Automaten eine totale Funktion (steht also in jeder Zelle der Übergangstabelle genau ein Element), so ist der Automat ein endlicher Automat mit vollständiger Übergangsfunktion

Sind endliche Automaten mit vollständiger Übergangsfunktion immer deterministisch?
Das Programm Exorciser bietet sehr gute Übungsmöglichkeiten für die Arbeit mit endlichen Automaten (Website)



# endliche Automaten mit $\epsilon$-Übergängen




Zu jedem endlichen Automaten mit $\epsilon$-Übergängen gibt es einen äquivalenten endlichen Automaten ohne $\epsilon$-Übergänge.



# Übung

Erstellen Sie endliche Automaten, die die folgenden Sprachen über dem Alphabet $\{a, b\}$ akzeptieren:

1. die Sprache aller Wörter, die nicht länger als 3 sind.
2. die Sprache aller Wörter, die mit 'ab' beginnen.
3. die Sprache aller Wörter, in denen die Kette 'aa' vorkommt.
4. die Sprache aller Wörter, die ungleich der Kette 'abb' sind.
5. die Sprache aller Wörter, die auf die Kette 'aa' enden.
6. die Sprache aller Wörter, in denen eine gerade Zahl von a's vorkommt.
7. die Sprache aller Wörter, in denen mindestens zwei a's vorkommen.

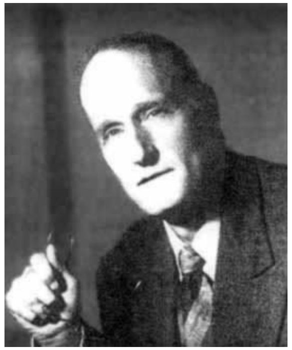


# reguläre Sprache

Gegeben ein Alphabet $\Sigma$.

- $\emptyset$ ist eine reguläre Sprache über dem Alphabet $\Sigma$.
- $\{\epsilon\}$ ist eine reguläre Sprache über dem Alphabet $\Sigma$.
- Für jedes $a \in \Sigma$ ist $\{a\}$ eine reguläre Sprache über dem Alphabet $\Sigma$.
- Wenn $A$ und $B$ reguläre Sprachen sind, dann ist auch $A \cup B$ eine reguläre Sprache über dem Alphabet $\Sigma$.
- Wenn $A$ und $B$ reguläre Sprachen sind, dann ist auch $A \circ B$ eine reguläre Sprache über dem Alphabet $\Sigma$.
- Wenn $A$ eine reguläre Sprachen ist, dann ist auch $A^*$ eine reguläre Sprache über dem Alphabet $\Sigma$.
- Nichts sonst ist eine reguläre Sprache über dem Alphabet $\Sigma$.

## Satz von Kleene



(Stephen C. Kleene, 1909 - 1994)

Jede Sprache, die von einem endlichen Automaten akzeptiert wird, ist regulär und jede reguläre Sprache wird von einem endlichen Automaten akzeptiert.



## Endliche Automaten akzeptieren reguläre Sprachen

**Theorem (Kleene)**

*Jede Sprache, die von einem endlichen Automaten akzeptiert wird ist regulär und jede reguläre Sprache wird von einem endlichen Automaten akzeptiert.*

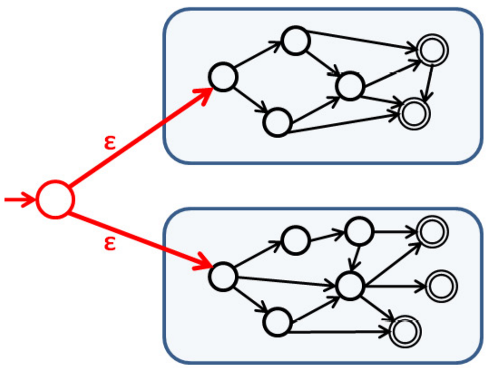
**Beweisidee (eine Richtung):** Zu jeder regulären Sprache gibt es einen endlichen Automaten, der diese akzeptiert:


$\emptyset$ $\{\varepsilon\}$ $\{a_i\}$



## Beweis des Satzes von Kleene (Fortsetzung)

Wenn $A$ und $B$ zwei reguläre Sprachen sind, die von den Automaten $A_A$ und $A_B$ akzeptiert werden, dann wird die reguläre Sprache $A \cup B$ von dem folgenden Automaten akzeptiert:

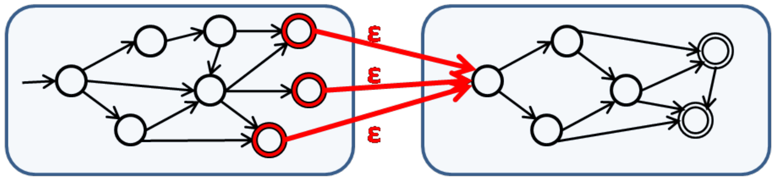



A ∪ B



## Beweis des Satzes von Kleene (Fortsetzung)

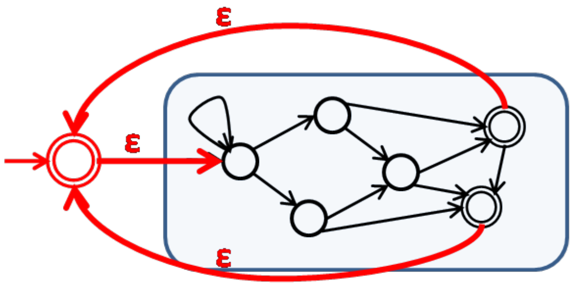
Wenn $A$ und $B$ zwei reguläre Sprachen sind, die von den Automaten $A_A$ und $A_B$ akzeptiert werden, dann wird die reguläre Sprache $A \circ B$ von dem folgenden Automaten akzeptiert:




A ∘ B

# Beweis des Satzes von Kleene (Fortsetzung)

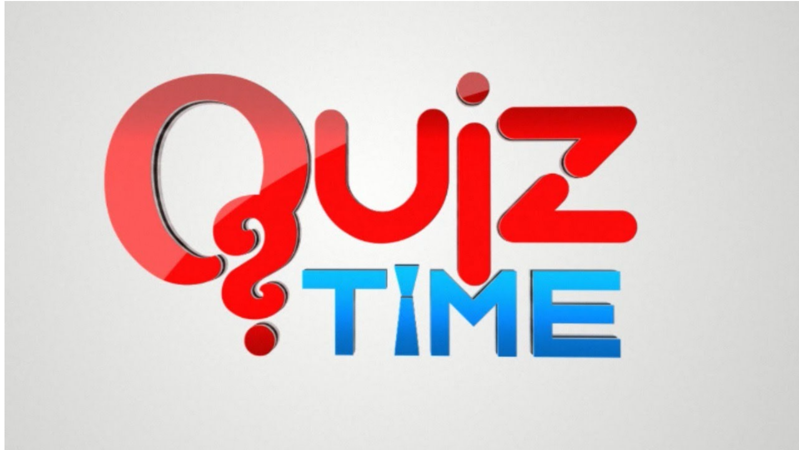
Wenn $A$ eine reguläre Sprache ist, die von dem Automaten $A_A$ akzeptiert wird, dann wird die reguläre Sprache $A^*$ von dem folgenden Automaten akzeptiert:




$A^*$



# Quiz-Time






# Formale Grammatik

## Definition

*Eine formale Grammatik ist ein 4-Tupel $G = (N, T, S, P)$ aus*

- *einem Alphabet von Terminalsymbolen $T$ (häufig auch $\Sigma$)*
- *einem Alphabet von Nichtterminalsymbolen $N$ mit $N \cap T = \emptyset$*
- *einem Startsymbol $S \in N$*
- *einer endlichen Menge von Regeln/Produktionen $P \subseteq \{(\alpha, \beta) \mid \alpha, \beta \in (N \cup T)^* \text{ und } \alpha \notin T^*\}$.*

*Für eine Regel $(\alpha, \beta)$ schreiben wir auch $\alpha \to \beta$.*
*Formale Grammatiken werden auch Typ0- oder allgemeine Regelgrammatiken genannt.*

$$\begin{array}{lllllllll} \text{S} & \to & \text{NP VP} & \text{VP} & \to & \text{V} & \text{NP} & \to & \text{D N} \\ \text{D} & \to & \text{the} & \text{N} & \to & \text{cat} & \text{V} & \to & \text{sleeps} \end{array}$$

Generiert: the cat sleeps



# Terminologie

$G = (\{\text{S,NP,VP,N,V,D,N,EN}\}, \{\text{the, cat, peter, chases}\}, S, P)$

$$P = \left\{ \begin{array}{lllllllll} \text{S} & \to & \text{NP VP} & \text{VP} & \to & \text{V NP} & \text{NP} & \to & \text{D N} \\ \text{NP} & \to & \text{EN} & \text{D} & \to & \text{the} & \text{N} & \to & \text{cat} \\ \text{EN} & \to & \text{peter} & \text{V} & \to & \text{chases} & & & \end{array} \right\}$$

"NP VP" ist in einem Schritt ableitbar aus $S$

"the cat chases peter" ist ableitbar aus $S$:

$$\begin{array}{llll} \text{S} & \vdash \text{NP VP} & \vdash \text{NP V NP} & \vdash \text{NP V EN} \\ & \vdash \text{NP V peter} & \vdash \text{NP chases peter} & \vdash \text{D N chases peter} \\ & \vdash \text{D cat chases peter} & \vdash \text{ the cat chases peter} & \end{array}$$

Die Menge aller aus dem Startsymbol $S$ ableitbarer Wörter (= Ketten aus Terminalsymbolen) ist die von der Grammatik $G$ erzeugte Sprache $L(G)$.

$$L(G) = \left\{ \begin{array}{cc} \text{the cat chases peter} & \text{peter chases the cat} \\ \text{peter chases peter} & \text{the cat chases the cat} \end{array} \right\}$$

Hinweis: für Terminalsymbole verwendet man in der Regel Klein- und für Nichtterminalsymbole Großbuchstaben.

# kontextfreie Grammatiken

Eine formale Grammatik, in der jede linke Regelseite aus genau einem Nichtterminalsymbol besteht, heißt kontextfrei.

Beispiel:
$G = (\{\text{S,NP,VP,N,V,D,N,EN}\}, \{\text{the, cat, peter, chases}\}, \text{S}, P)$

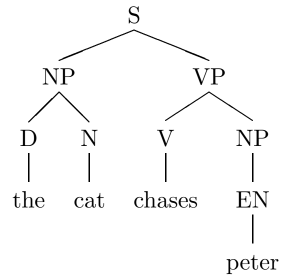
$$P = \left\{ \begin{array}{rcl rcl rcl} \text{S} & \rightarrow & \text{NP VP} & \text{VP} & \rightarrow & \text{V NP} & \text{NP} & \rightarrow & \text{D N} \\ \text{NP} & \rightarrow & \text{EN} & \text{D} & \rightarrow & \text{the} & \text{N} & \rightarrow & \text{cat} \\ \text{EN} & \rightarrow & \text{peter} & \text{V} & \rightarrow & \text{chases} & & & \end{array} \right\}$$



# Linksableitung (kontextfreie Grammatiken)

Gegeben eine kontextfreie Grammatik G. Eine Ableitung bei der stets das am weitesten links stehende nichtterminale Symbol ersetzt wird, heißt Linksableitung

$$\begin{array}{llll} \text{S} & \vdash \text{NP VP} & \vdash \text{D N VP} & \vdash \text{the N VP} \\ & \vdash \text{the cat VP} & \vdash \text{the cat V NP} & \vdash \text{the cat chases NP} \\ & \vdash \text{the cat chases EN} & \vdash \text{the cat chases peter} & \end{array}$$




Zu jeder Linksableitung gibt es genau einen Ableitungsbaum und zu jedem Ableitungsbaum gibt es genau eine Linksableitung.



# Chomskyhierarchie

- Wenn man die Form der Regeln einschränkt, erhält man Teilmengen der Menge aller durch eine Grammatik erzeugten Sprachen.
- Die Chomskyhierarchie ist eine Hierarchie über die Regelbedingungen (den verschiedenen Sprachklassen entsprechen Einschränkungen über die rechten und linken Regelseiten).
- Die Chomskyhierarchie reflektiert eine spezielle Form der Komplexität, andere Kriterien sind denkbar und führen zu anderen Hierarchien.
- Die Sprachklassen der Chomskyhierarchie sind in der Informatik intensiv untersucht worden (Berechnungskomplexität, effektive Parser).
- Für Linguisten ist die Chomskyhierarchie besonders interessant, da sie die Form der Regeln zentral stellt, und somit Aussagen über Grammatikformalismen zulässt.



# Noam Chomsky

Noam Chomsky
(∗ 7.12.1928, Philadelphia)
Noam Chomsky, *Three Models for the Description of Language*, IRE Transactions on Information Theory (1956).

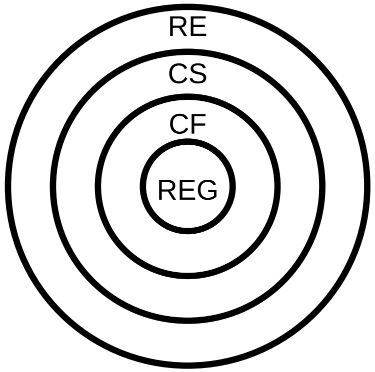
## Chomsky-Hierarchie & Automaten (grober Überblick)

| *Sprache* | *Automat* | *Grammatik* | *Erkennung* | *Abhängigkeit* |
|---|---|---|---|---|
| rekursiv aufzählbar | Turing Maschine | unbeschränkt<br>Baa → ε | unentscheidbar | beliebig |
| kontext-sensitiv | linear gebunden | kontext-sensitiv<br>γAδ → γβδ | NP-vollständig | überkreuzt |
| kontext-frei | Kellerautomat (Stapel) | kontextfrei<br>C → bABa | polynomiell | eingebettet |
| regulär | endlicher Automat | regulär<br>A → bA | linear | strikt lokal |



## Chomskyhierarchie: Hauptsatz

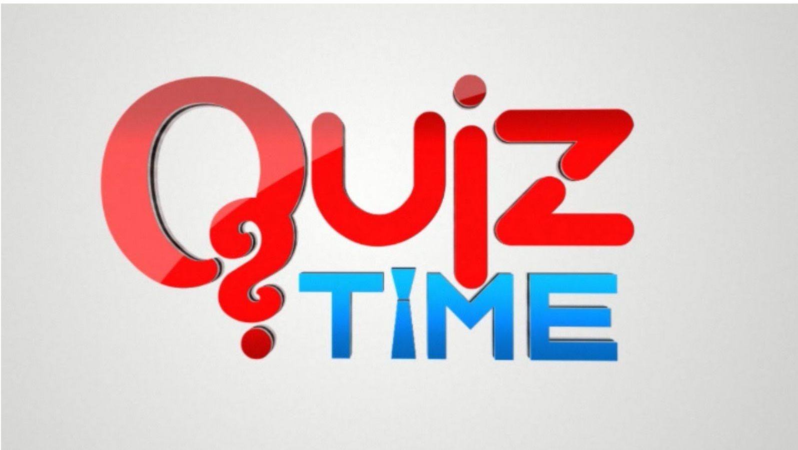
**REG ⊂ CF ⊂ CS ⊂ RE**




REG: reguläre Sprachen, CF: kontextfreie Sprachen, CS: kontextsensitive Sprachen, RE: rekursiv-aufzählbare Sprachen



## Quiz-Time






# Mathematische Grundlagen der Computerlinguistik

## Ordnungsrelationen

Dozentin: Wiebke Petersen

4. Foliensatz

## strikte / schwache Ordnungen

Eine Ordnung $R$ einer Menge $A$ ist eine binäre Relation $R \subseteq A \times A$.
Man unterscheidet zwischen strikten und schwachen Ordnungen:

### Eine binäre Relation ist eine schwache Ordnung, gdw. sie

- transitiv,
- reflexiv und
- anti-symmetrisch

ist.

### Eine binäre Relation ist eine strikte Ordnung, gdw. sie

- transitiv,
- irreflexiv und
- asymmetrisch

ist.

Strikte Ordnungen werden auch starke Ordnungen genannt.



## korrespondierende Ordnungen

Eine schwache Ordnung $R \subseteq A \times A$ und eine strikte Ordnung $S$ korrespondieren zueinander gdw.

$$R = S \cup id_A$$

Beispiele: Sei $A = \{a, b, c, d\}$

- $R_1 = \{\langle a, b\rangle, \langle a, c\rangle, \langle a, d\rangle, \langle b, c\rangle, \langle a, a\rangle, \langle b, b\rangle, \langle c, c\rangle, \langle d, d\rangle\}$
- $R_2 = \{\langle b, a\rangle, \langle c, b\rangle, \langle c, a\rangle, \langle b, b\rangle, \langle a, a\rangle, \langle c, c\rangle, \langle d, d\rangle\}$
- $R_3 = \{\langle d, c\rangle, \langle d, b\rangle, \langle d, a\rangle, \langle c, b\rangle, \langle c, a\rangle, \langle b, a\rangle, \langle a, a\rangle, \langle b, b\rangle, \langle c, c\rangle, \langle d, d\rangle\}$

korrespondierende strikte Ordnungen:

- $S_1 = \{\langle a, b\rangle, \langle a, c\rangle, \langle a, d\rangle, \langle b, c\rangle\}$
- $S_2 = \{\langle b, a\rangle, \langle c, b\rangle, \langle c, a\rangle\}$
- $S_3 = \{\langle d, c\rangle, \langle d, b\rangle, \langle d, a\rangle, \langle c, b\rangle, \langle c, a\rangle, \langle b, a\rangle\}$



## geordnete Mengen

Eine geordnete Menge ist ein Paar $(M, R)$, bestehend aus einer Menge $M$ und einer Ordnung $R$ von $M$.

Beispiele:

- $(\mathcal{POT}(M), \subseteq)$ ist eine schwach geordnete Menge.
  $(\mathcal{POT}(M), \subset)$ ist die korrespondierende strikt geordnete Menge.
- $(\mathbb{N}, \leq)$ ist eine schwach geordnete Menge.
  $(\mathbb{N}, <)$ ist die korrespondierende strikt geordnete Menge.



## Terminologie

Sei $(M, R)$ eine (strikt oder schwach) geordnete Menge.

- $a$ ist ein Vorgänger von $b$ gdw. $R(a, b)$.
- $a$ ist ein Nachfolger von $b$ gdw. $R(b, a)$.
- $a$ ist ein unmittelbarer Vorgänger (oder unterer Nachbar) von $b$ gdw.
  - $a \neq b$,
  - $R(a, b)$, und
  - es gibt kein $c \in M$ mit $c \notin \{a, b\}$ so dass $R(a, c)$ und $R(c, b)$.
- $a$ ist ein unmittelbarer Nachfolger (oder oberer Nachbar) von $b$ gdw. $b$ ein unmittelbarer Vorgänger von $a$ ist.

Wenn $a$ ein unmittelbarer Vorgänger von $b$ ist, dann schreibt man häufig $a \prec b$.

# Hassediagramm

## Konstruktion

Eine endliche geordnete Mengen $(M, R)$ kann durch ein Hassediagramm veranschaulicht werden; dieses erhält man, indem man für jedes Element von $M$ einen Punkt zeichnet und zwar so, daß $a$ unterhalb von $b$ liegt, wenn $a \neq b$ und $(a, b) \in R$.

Zwei Punkte $a$ und $b$ werden mit einer Linie verbunden, wenn $a \prec b$.

Übung: Zeichnen sie die folgenden Hasse-Diagramme

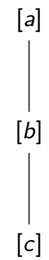
Hasse-Diagramm von $(\{a, b, c\}, R_2)$ mit $R_2 = \{\langle b, a\rangle, \langle c, b\rangle, \langle c, a\rangle, \langle b, b\rangle, \langle a, a\rangle, \langle c, c\rangle\}$

Hasse-Diagramm von $(\mathcal{POT}(\{1, 2, 3\}), \subseteq)$

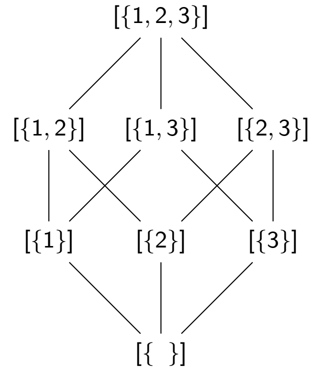


# Beispiele

Hasse-Diagramm von $(\{a, b, c\}, R_2)$ mit $R_2 = \{\langle b, a\rangle, \langle c, b\rangle, \langle c, a\rangle, \langle b, b\rangle, \langle a, a\rangle, \langle c, c\rangle\}$




Hasse-Diagramm von $(\mathcal{POT}(\{1, 2, 3\}), \subseteq)$

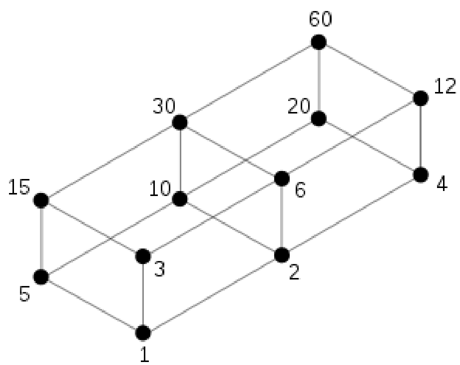





# Hasse-Diagramme: Beispiel Teilbarkeit

Sei $M = \{x \in \mathbb{N} \mid 60 \text{ ist durch } x \text{ ohne Rest teilbar}\}$, und $R = \{\langle x, y\rangle \in M \times M \mid y \text{ ist durch } x \text{ ohne Rest teilbar}\}$.

Hassediagramm der geordneten Menge $(M, R)$:






# Übung

Zeichnen sie ein Hasse-Diagramm zur geordneten Menge $M = (\{\{1, 2, 3, 4, 5\}, \{1, 2, 3, 5\}, \{1, 3, 4\}, \{2, 4, 5\}, \{1, 2, 3\}, \{1, 3\}, \{2, 4\}, \{1, 5\}, \{1\}, \{3\}, \{4\}, \{5\}, \{\}\}, \subseteq)$.

## totale/partielle Ordnung

Eine binäre Ordnungsrelation ist eine totale Ordnung, gdw. sie konnex ist.

Eine binäre Relation $R \subseteq M \times M$ ist konnex (bzw. linear) gdw. für alle $x, y \in M$ mit $x \neq y$ gilt: $\langle x, y \rangle \in R$ oder $\langle y, x \rangle \in R$.

- Das Hassediagramm einer total geordneten, endlichen Menge bildet eine Linie. Kein Element hat mehr als einen oberen oder unteren Nachbarn.
- Totale Ordnungen werden auch lineare Ordnungen genannt.
- In Abgrenzung zu totalen Ordnungen werden allgemeine Ordnungen auch partielle Ordnungen (oder Halbordnungen) genannt. Im Englischen spricht man von 'poset' (partially ordered set).



## minimale und maximale Elemente

Sei $R \subseteq A \times A$ eine Ordnung (strikt oder schwach).

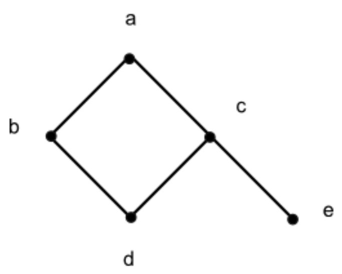
- Ein Element $x \in A$ ist minimal gdw. es kein $y \neq x$ gibt, das Vorgänger von $x$ ist.
- Ein Element $x \in A$ ist maximal gdw. es kein $y \neq x$ gibt, das Nachfolger von $x$ ist.
- $x \in A$ ist das Minimum von $A$, wenn $x$ Vorgänger jedes anderen Elements von $A$ ist (für alle $y \in A$ mit $x \neq y$ gilt $xRy$).
- $x \in A$ ist das Maximum von $A$, wenn $x$ Nachfolger jedes anderen Elements von $A$ ist (für alle $y \in A$ mit $x \neq y$ gilt $yRx$).

Hinweise:

- eine total geordnete Menge kann höchstens ein minimales und höchstens ein maximales Element haben.
- eine partiell geordnete Menge kann beliebig viele minimale und maximale Elemente aber höchstens ein Minimum und höchstens ein Maximum haben.



## Beispiel




- $a$ ist das einzige maximale Element und somit das Maximum der geordneten Menge.
- $d$ und $e$ sind die minimalen Elemente der geordneten Menge.
- die geordnete Menge hat kein Minimum,



## Vergleichbarkeit / Kette / Antikette

Sei $(M, R)$ eine geordnete Menge und seien $a$ und $b$ Elemente von $M$. $a$ und $b$ heißen vergleichbar, falls $aRb$ oder $bRa$; sonst unvergleichbar. Eine Teilmenge $K$ von $M$ heißt Kette, g.d.w. für beliebige $a, b \in K$ mit $a \neq b$ gilt, daß sie vergleichbar sind. Eine Teilmenge $A$ von $M$ heißt Antikette, g.d.w. für beliebige $a, b \in A$ mit $a \neq b$ gilt, daß sie unvergleichbar sind.

### Satz von Dilworth

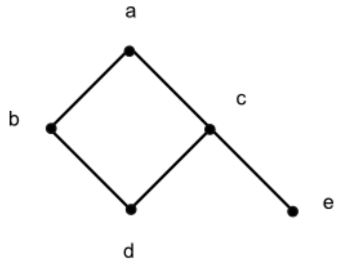
Für eine geordnete endliche Menge $(M, R)$ gilt: Die maximale Anzahl von Elementen in einer Antikette von $(M, R)$ ist gleich der kleinsten Anzahl von Ketten von $(M, R)$, die man für eine Partition von $M$ benötigt.

### Höhe / Breite

Die Höhe einer endlichen geordneten Menge $(M, R)$ ist gleich der maximalen Anzahl von Elementen einer Kette von $(M, R)$.

Die Breite einer endlichen geordneten Menge $(M, R)$ ist gleich der maximalen Anzahl von Elementen einer Antikette von $(M, R)$.

## Beispiel




- Die Elemente $a$ und $b$ sind vergleichbar.
- $d$ und $e$ sind unvergleichbar.
- $\{a, b, d\}$ ist eine Kette der geordneten Menge.
- $\{b, c\}$ ist Antikette der geordneten Menge.
- Die geordnete Menge hat die Höhe 3 und die Breite 2.
- Die Ketten $\{a, b, d\}$ und $\{c, e\}$ bilden eine minimale Partition in Ketten der geordneten Menge.



## Intervall / Ideal / Filter

Sei $(M, \trianglelefteq)$ eine geordnete Menge und $a, b \in M$ mit $a \trianglelefteq b$:

Intervall: $[a, b] := \{x \in M \mid a \trianglelefteq x \trianglelefteq b \text{ oder } x = a \text{ oder } x = b\}$

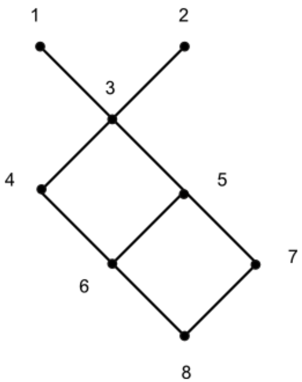
Hauptideal: $(b] := \{x \in M \mid x \trianglelefteq b \text{ oder } x = b\}$

Hauptfilter: $[a) := \{x \in M \mid a \trianglelefteq x \text{ oder } x = a\}$

Hinweis: $\trianglelefteq$ steht hier für eine beliebige Ordnungsrelation. Man hätte auch "Sei $(M, R)$ eine geordnete Menge: ..." schreiben können.



## Beispiel




- $[6, 1] = \{1, 3, 4, 5, 6\}$ (Intervall von 6 bis 1)
- $(4] = \{4, 6, 8\}$ (Hauptideal von 4)
- $[6) = \{1, 2, 3, 4, 5, 6\}$ (Hauptfilter von 6).



## Quasiordnung

Der Begriff der Quasiordnung ist schwächer als der der Ordnung:

### Definition

Eine binäre Relation $R \subseteq M \times M$ ist eine Quasiordnung (oder Präordnung), wenn $R$

- reflexiv und
- transitiv ist.

Beispiel:

- Die Relation $\leq_{abs}$, die die ganzen Zahlen nach ihrem Betrag ordnet ist eine Quasiordnung aber keine Ordnung (beachte, dass $-3 \leq_{abs} 3$ und $3 \leq_{abs} -3$ aber $-3 \neq 3$).

# Zusammenfassung: Ordnungen

## schwache Ordnungen

| | transitiv | reflexiv | anti-symmetrisch | linear/total |
|---|---|---|---|---|
| Quasiordnung | $\star$ | $\star$ | | |
| partielle Ordnung | $\star$ | $\star$ | $\star$ | |
| totale Ordnung | $\star$ | $\star$ | $\star$ | $\star$ |

Bemerkung: (Schwache) lineare Ordnungsrelationen werden häufig mit $\leq$, bzw. partielle Ordnungsrelationen mit $\subseteq$ bezeichnet, auch wenn es sich bei der gegebenen Ordnung weder um eine numerische Größenordnung noch um die Mengeninklusion handelt.



# Zusammenfassung: Ordnungen

## strikte Ordnungen

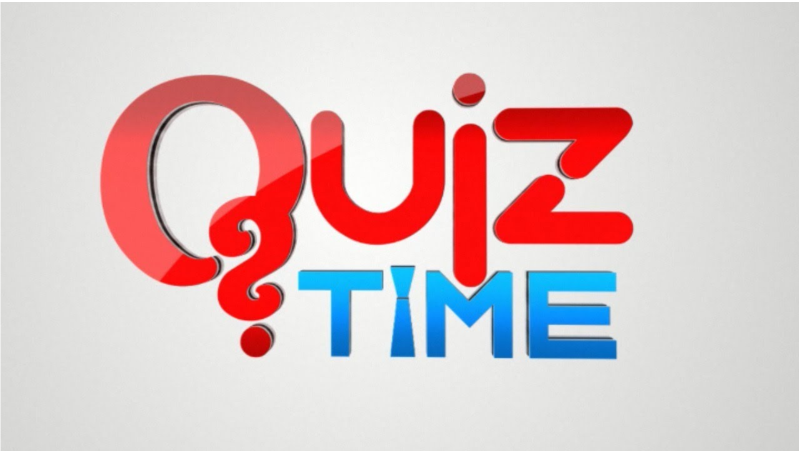
| | transitiv | irreflexiv | asymmetrisch | linear/total |
|---|---|---|---|---|
| strikte partielle Ordnung | $\star$ | $\star$ | $\star$ | |
| strikte totale Ordnung | $\star$ | $\star$ | $\star$ | $\star$ |

Bemerkung: Strikte Ordnungsrelationen werden häufig mit $<$, bzw. mit $\subset$ bezeichnet.

Man könnte strikte Ordnungen äquivalent auch als transitive, irreflexive und antisymmetrische Relationen definieren, da eine Relation, die irreflexiv und antisymmetrisch ist, immer asymmetrisch ist.



# Quiz-Time






# Ordnungserhaltende Abbildungen

## Definition

Seien $(M, \trianglelefteq)$ und $(M', \trianglelefteq')$ zwei geordnete Mengen. Eine Abbildung (Funktion) $f : M \to M'$ heißt ordnungserhaltend, wenn für alle $x, y \in M$ gilt:

$$\text{wenn } x \trianglelefteq y, \text{ dann } f(x) \trianglelefteq' f(y)$$

Eine ordnungserhaltende Abbildung ist ein Ordnungshomomorphismus.

Beispiele:

- $f : \mathbb{N}_0 \to \mathbb{N}_0$ mit $f(x) = 2x$ ist eine ordnungserhaltende Abbildung von $(\mathbb{N}_0, \leq)$ nach $(\mathbb{N}_0, \leq)$.
- $f : \mathbb{N}_0 \to \mathbb{N}_0$ mit $f(x) = x^2$ ist eine ordnungserhaltende Abbildung von $(\mathbb{N}_0, \leq)$ nach $(\mathbb{N}_0, \leq)$.
- $f : \mathbb{Z} \to \mathbb{Z}$ mit $f(x) = x^2$ ist keine ordnungserhaltende Abbildung von $(\mathbb{Z}, \leq)$ nach $(\mathbb{Z}, \leq)$.
- Sei $M$ eine endliche Menge. $f : \mathcal{POT}(M) \to \mathbb{N}_0$ mit $f(A) = |A|$ ist eine ordnungserhaltende Abbildung von $(\mathcal{POT}(M), \subseteq)$ nach $(\mathbb{N}_0, \leq)$ .

## Terminologie und Anmerkungen

- Ordnungserhaltende Abbildungen nennt man auch isotone Abbildungen.
- Abbildungen für die aus $x \trianglelefteq y$ folgt, dass $f(y) \trianglelefteq' f(x)$ gilt, heißen antiton.
- Eine Abbildung ist monoton, wenn sie isoton oder antiton ist.
- **Vorsicht:** Häufig wird der Begriff 'monoton' auch nur für isotone Abbildungen verwendet.



## Ordnungsreflektierende Abbildungen

### Definition

Seien $(M, \trianglelefteq)$ und $(M', \trianglelefteq')$ zwei geordnete Mengen. Eine Abbildung (Funktion) $f : M \to M'$ heißt ordnungsreflektierend, wenn für alle $x, y \in M$ gilt:

$$\text{wenn } f(x) \trianglelefteq' f(y), \text{ dann } x \trianglelefteq y$$

Beispiele:

- $f : \mathbb{N}_0 \to \mathbb{N}_0$ mit $f(x) = 2x$ ist eine ordnungsreflektierende Abbildung von $(\mathbb{N}_0, \leq)$ nach $(\mathbb{N}_0, \leq)$.
- Sei $M$ eine endliche Menge. $f : \mathcal{POT}(M) \to \mathbb{N}_0$ mit $f(A) = |A|$ ist keine ordnungsreflektierende Abbildung von $(\mathcal{POT}(M), \subseteq)$ nach $(\mathbb{N}_0, \leq)$ .

**Satz:** Ordnungsreflektierende Abbildungen sind immer injektiv.



## Ordnungseinbettung

### Definition

Seien $(M, \trianglelefteq)$ und $(M', \trianglelefteq')$ zwei geordnete Mengen. Eine Abbildung (Funktion) $f : M \to M'$ heißt Ordnungseinbettung, wenn sie ordnungserhaltend und ordnungsreflektierend ist, wenn also folgendes gilt:

$$x \trianglelefteq y \Leftrightarrow f(x) \trianglelefteq' f(y)$$

Ein Ordnungsisomorphismus ist eine bijektive Ordnungseinbettung. Gibt es einen Ordnungsisomorphismus zwischen $(M, \trianglelefteq)$ und $(M', \trianglelefteq')$, so sagt man dass die beiden geordneten Mengen (ordnungs-)isomorph sind und schreibt $(M, \trianglelefteq) \cong (M', \trianglelefteq')$.

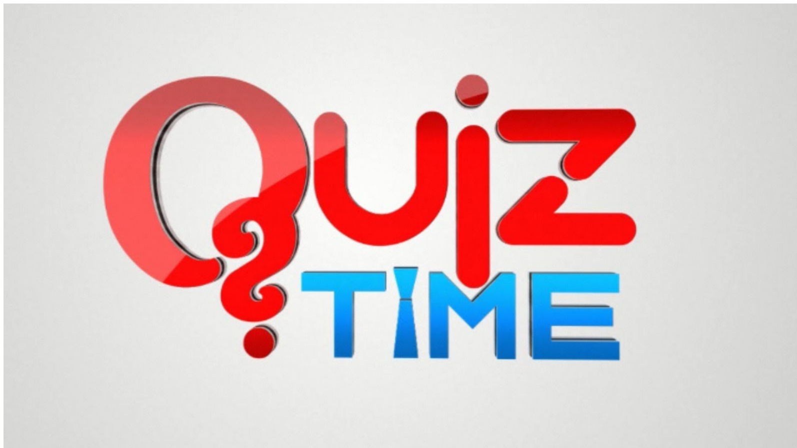
Ein Ordnungsisomorphismus von $(M, R)$ in sich selbst wird auch Ordnungsautomorphismus genannt.

Beispiele:

- $f : \mathbb{Z} \to \mathbb{Z}$ mit $f(x) = -x$ ist ein Ordnungsisomorphismus von $(\mathbb{Z}, \leq)$ nach $(\mathbb{Z}, \geq)$.
- $f : \mathbb{R} \to \mathbb{R}$ mit $f(x) = \frac{x}{2}$ ist ein Ordnungsautomorphismus auf $(\mathbb{R}, \leq)$.



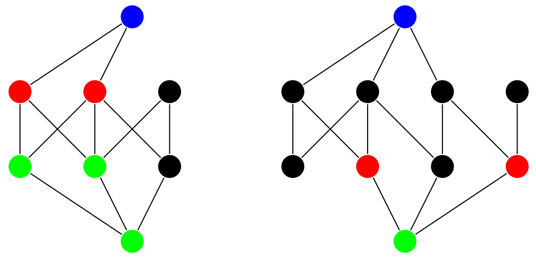
## Quiz-Time



## obere / untere Schranke

Sei $(M, \trianglelefteq)$ eine geordnete Menge und $K$ eine Teilmenge von $M$. Ein Element $x$ von $M$ ist

- eine obere Schranke von $K$, g.d.w. für alle $y \in K : y \trianglelefteq x$ oder $y = x$;
- eine untere Schranke von $K$, g.d.w. für alle $y \in K : x \trianglelefteq y$ oder $y = x$.



Die Abbildungen zeigen die Hassediagramme zweier geordneter Mengen. Die rot markierten Elementen haben die blau markierten Elemente als obere und die grün markierten als untere Schranken.



## kleinste obere / größte untere Schranke

$x$ heißt kleinste obere Schranke oder Supremum von $K$ in $M$, wenn $x$ eine obere Schranke von $K$ ist und für jede obere Schranke $y \in M$ von $K$ mit $x \neq y$ die Ungleichung $x \trianglelefteq y$ gilt. Wir schreiben $\sup K$ oder $\bigvee K$ für das Supremum von $K$ (lese $\vee$ als 'join').

$x$ heißt größte untere Schranke oder Infimum von $K$ in $M$, wenn $x$ eine untere Schranke von $K$ ist und für jede untere Schranke $y \in M$ von $K$ mit $x \neq y$ die Ungleichung $y \trianglelefteq x$ gilt. Wir schreiben $\inf K$ oder $\bigwedge K$ für das Infimum von $K$ (lese $\wedge$ als 'meet').

Wir schreiben $x \vee y$ statt $\bigvee\{x, y\}$ und $x \wedge y$ statt $\bigwedge\{x, y\}$.

Die Beispiele der vorangegangenen Folie zeigen, daß es geordnete Mengen $M$ gibt, für die nicht jede Teilmenge $K \subseteq M$ ein Supremum oder Infimum hat.

Das Infimum ist also das Maximum aller unteren Schranken und das Supremum ist das Minimum aller oberen Schranken.



## Beispiele

- Für die linear geordnete Menge $(\mathbb{R}, \leq)$ gilt: $\sup[1, 4] = 4$ und $\inf[1, 4] = 1$.
- Für die partiell geordnete Menge $(\mathcal{POT}(M), \subseteq)$ mit $M = \{1, 2, 3, 4\}$ ist das Supremum von $K = \{\{1, 2\}, \{2, 4\}, \{1\}\}$ die Vereinigung aller Elemente von $K$, also $\sup K = \{1, 2, 4\}$.
  Das Infimum von $K$ ist der Durchschnitt aller Elemente von $K$, also $\inf K = \emptyset$.



## Verbände

**Verband: ordnungstheoretische Definition**

Eine geordnete Menge $(V, \leq)$ ist ein Verband, g.d.w. zu je zwei Elementen $x$ und $y$ aus $V$ auch das Supremum von $x$ und $y$ und das Infimum von $x$ und $y$ Elemente von $V$ sind.

**vollständiger Verband**

Ein Verband $(V, \leq)$ ist ein vollständiger Verband, falls für alle $K \subseteq V$ gilt, daß $\sup K \in V$ und $\inf K \in V$.

Jeder vollständige Verband hat ein größtes Element $\sup V$, das Einselement ($1_V$) genannt, und ein kleinstes Element $\inf V$, das Nullement ($0_V$) genannt.

Die oberen Nachbarn des Nullelements nennt man die Atome und die unteren Nachbarn des Einselements die Koatome des Verbands.

## Bemerkungen

- Jeder nichtleere endliche Verband ist vollständig.
- Da $\emptyset \subseteq V$ und $\inf \emptyset = 1_V$ und $\sup \emptyset = 0_V$ gilt, gibt es keinen vollständigen Verband mit leerer Menge $V$.

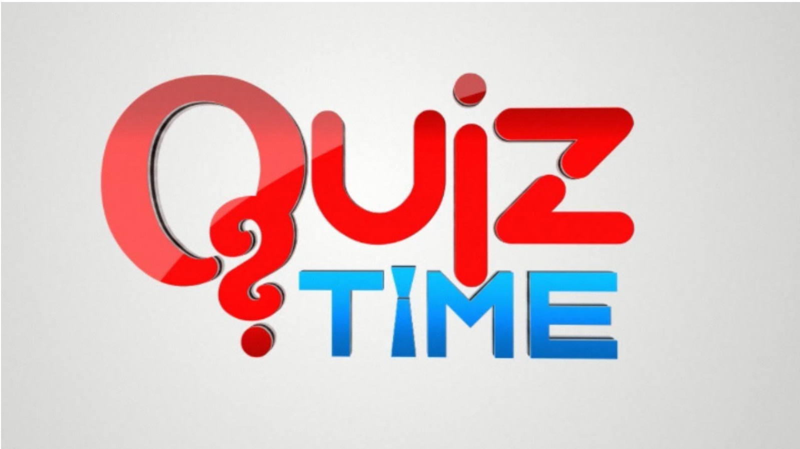


## Beispiele

- $(\mathcal{POT}(M), \subseteq)$ ist ein vollständiger Verband, $\vee$ entspricht $\cup$ und $\wedge$ entspricht $\cap$.
- $([2,5], \leq)$ ist ein vollständiger Verband.
- $(\mathbb{R}, \leq)$ ist ein Verband, aber nicht vollständig.
- $\big(\{\{1,2\},\{2,4\},\{1\}\}, \subseteq\big)$ ist kein Verband.



## Quiz-Time






# Mathematische Grundlagen der Computerlinguistik

## Algebren

Dozentin: Wiebke Petersen

5. Foliensatz

## Algebren (algebraische Strukturen)

Eine Algebra **A** ist eine Menge $A$ zusammen mit einer oder mehreren n-stelligen Operationen (Verknüpfungen) $f_i$.

In diesem Kurs beschränken wir uns auf Algebren mit ein oder zwei binären Operationen.

Die Operationen einer Algebra müssen die folgenden Axiome erfüllen:

Abgeschlossenheit: $A$ ist unter der Operation $\otimes$ abgeschlossen, d.h. für beliebige $a, b \in A$ gibt es ein Element $c \in A$, sodass $a \otimes b = c$.

Eindeutigkeit: Wenn $a = a'$ und $b = b'$, dann gilt $a \otimes b = a' \otimes b'$.

An was erinnern Sie die beiden Axiome?

Alternative Definition: Eine Algebra **A** ist eine Menge $A$ zusammen mit einer oder mehreren n-stelligen Funktionen $f_i : A^n \to A$.



## Eigenschaften von Operationen

### Assoziativgesetz

Eine Operation $\otimes$ auf $A$ ist assoziativ, g.d.w. für alle $a, b, c \in A$ gilt:

$$a \otimes (b \otimes c) = (a \otimes b) \otimes c$$

### Kommutativgesetz

Eine Operation $\otimes$ auf $A$ ist kommutativ, g.d.w. für alle $a, b \in A$ gilt:

$$a \otimes b = b \otimes a$$

### Idempotenzgesetz

Eine Operation $\otimes$ auf $A$ ist idempotent, g.d.w. für alle $a \in A$ gilt:

$$a \otimes a = a$$

### Distributivgesetz

Für zwei Operationen $\oplus$ und $\otimes$ auf $A$ distributiert $\oplus$ über $\otimes$, g.d.w. für alle $a, b, c \in A$ gilt:

$$a \oplus (b \otimes c) = (a \oplus b) \otimes (a \oplus c)$$



## neutrale und inverse Elemente

### neutrales Element

Gegeben eine Operation $\oplus$ auf $A$. Ein Element $e \in A$ ist das neutrale Element von $\oplus$, g.d.w. für alle $a$ in $A$ gilt:

$$e \oplus a = a \oplus e = a.$$

### inverses Element

Gegeben eine Operation $\oplus$ auf $A$ mit neutralem Element $e$. Ein Element $a^{-1} \in A$ ist das inverse Element eines Elements $a \in A$, g.d.w.:

$$a^{-1} \oplus a = a \oplus a^{-1} = e$$



## Bsp: Drehungen eines gleichseitigen Dreiecks $((\triangle_D, \circ))$

Grundmenge: $\triangle_D = \{id, \curvearrowright, \curvearrowleft\}$
(id: 0-Drehung; $\curvearrowright$: 120-Drehung nach rechts; $\curvearrowleft$: 120-Drehung nach links)
Operation: $\circ$ Hintereinanderausführung der Drehungen.

| $\circ$ | **id** | $\curvearrowright$ | $\curvearrowleft$ |
|---|---|---|---|
| **id** | id | $\curvearrowright$ | $\curvearrowleft$ |
| $\curvearrowright$ | $\curvearrowright$ | $\curvearrowleft$ | id |
| $\curvearrowleft$ | $\curvearrowleft$ | id | $\curvearrowright$ |

- neutrales Element: id
- inverse Elemente: $\text{id}^{-1} = \text{id}$; $\curvearrowright^{-1} = \curvearrowleft$; $\curvearrowleft^{-1} = \curvearrowright$
- Eigenschaften von $\circ$: assoziativ, kommutativ

## Bsp: Drehungen und vertikale Spiegelung eines gleichseitigen Dreiecks

Grundmenge: $\{id, \curvearrowright, \curvearrowleft, \leftrightarrows\}$
(id: 0-Drehung; $\curvearrowright$: 120-Drehung nach rechts; $\curvearrowleft$: 120-Drehung nach links, $\leftrightarrows$: vertikale Spiegelung)
Operation: $\circ$ Hintereinanderausführung der Drehungen und Spiegelungen.

Diese Struktur bildet keine Algebra, da u.a. $\leftrightarrows \circ \curvearrowleft$ kein Element der Grundmenge ist (Verletzung der Abgeschlossenheit).

Wenn man alle drei Spiegelungen entlang aller drei Spiegelachsen hinzunimmt, erhält man wieder eine Algebra.



## Beispiel: Restklassen modulo 3 $((\mathbb{N} \mod 3, \oplus_3))$

Grundmenge: $\{[0], [1], [2]\}$
Operation: $\oplus_3$: Summe modulo 3

| $\oplus_3$ | [0] | [1] | [2] |
|---|---|---|---|
| [0] | [0] | [1] | [2] |
| [1] | [1] | [2] | [0] |
| [2] | [2] | [0] | [1] |

- neutrales Element: [0]
- inverse Elemente: $[0]^{-1} = [0]$; $[1]^{-1} = [2]$; $[2]^{-1} = [1]$
- Eigenschaften von $\oplus_3$: assoziativ, kommutativ

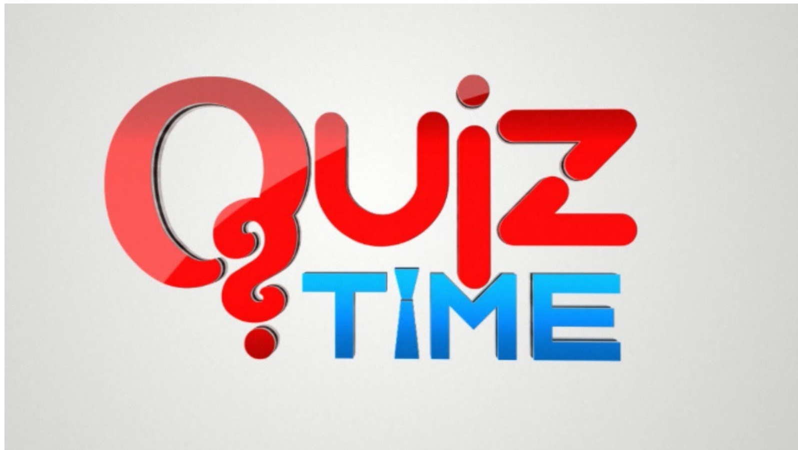


## weitere Beispiele für Algebren

- $(\mathbb{N}_0, +)$
- $(\mathbb{Z}, +, \cdot)$
- $(\mathbb{R}, +, \cdot)$
- $(\mathcal{POT}(M), \cap, \cup)$
- $(\Sigma^*, \circ)$



## Quiz-Time



# Morphismen

## Morphismus

Ein Morphismus ($\varphi : \mathbf{A} \to \mathbf{B}$) von einer Algebra $\mathbf{A}$ in eine Algebra $\mathbf{B}$ ist eine Abbildung, die zum einen eine Funktion von der Menge der ersten in die Menge der zweiten Algebra definiert ($F : A \to B$), und zum anderen die Operationen der ersten Algebra auf die zweite Algebra projiziert (hierzu müssen beide Algebren gleichviele Operationen gleicher Stelligkeit haben).

## Homomorphismus

Gegeben zwei Algebren $\mathbf{A} = (A, \oplus, \otimes)$ und $\mathbf{B} = (B, \star, \circ)$. Ein Morphismus $\varphi : \mathbf{A} \to \mathbf{B}$ ist ein Homomorphismus, g.d.w. für alle $x, y$ in $A$ gilt:

$$\varphi(x) \star \varphi(y) = \varphi(x \oplus y) \text{ und}$$
$$\varphi(x) \circ \varphi(y) = \varphi(x \otimes y)$$



## Isomorphismus

Gegeben zwei Algebren $\mathbf{A} = (A, \oplus, \otimes)$ und $\mathbf{B} = (B, \star, \circ)$. Ein Morphismus $\varphi : \mathbf{A} \to \mathbf{B}$ ist ein Isomorphismus, g.d.w. $\varphi : A \to B$ bijektiv ist und wenn für alle $x, y$ in $A$ gilt:

$$\varphi(x) \star \varphi(y) = \varphi(x \oplus y) \text{ und}$$
$$\varphi(x) \circ \varphi(y) = \varphi(x \otimes y)$$

(Isomorphismen sind also bijektive Homomorphismen)

Zwei Algebren sind isomorph, wenn es einen Isomorphismus zwischen ihnen gibt.

## Automorphismus

Ein Automorphismus einer Algebra $\mathbf{A}$ ist ein Isomorphismus $\varphi : \mathbf{A} \to \mathbf{A}$.



# Beispiele

- $\varphi : (\mathbb{N}_0, +) \to (\mathbb{N} \mod 3, \oplus_3)$ mit $\varphi(n) = n \mod 3$ ist ein Homomorphismus, aber kein Isomorphismus
- $\varphi : (\mathbb{N}_0, +) \to (\{a\}^*, \circ)$ mit $\varphi(n) = a^n$ ist ein Isomorphismus.
- $\varphi : (\triangle_D, \circ) \to (\mathbb{N} \mod 3, \oplus_3)$ mit $\varphi(\mathrm{id}) = [0]$, $\varphi(\curvearrowright) = [1]$, $\varphi(\curvearrowleft) = [2]$ ist ein Isomorphismus.



# Semigruppe, Monoid, Gruppe

## Semigruppe

Eine Semigruppe (Halbgruppe) $\mathbf{G} = (G, \otimes)$ ist eine Algebra, bestehend aus einer Menge $G$ und einer binären Operation $\otimes$, die folgende Bedingungen erfüllt:

G1 $\otimes$ ist assoziativ

## Monoid

Ein Monoid $\mathbf{G} = (G, \otimes)$ ist eine Algebra mit:

G1 $\otimes$ ist assoziativ

G2 $G$ enthält ein neutrales Element

## Gruppe

Eine Gruppe $\mathbf{G} = (G, \otimes)$ ist eine Algebra mit:

G1 $\otimes$ ist assoziativ

G2 $G$ enthält ein neutrales Element

G3 jedes Element aus $G$ hat ein inverses Element in $G$.

# Beispiele

- $(\mathbb{N}, +)$ ist eine Semigruppe
- $(\mathbb{N}_0, +)$ ist ein Monoid
- $(\mathbb{Z}, +)$ ist eine Gruppe
- $(\mathcal{POT}(M), \cup)$ ist ein Monoid
- $(\Sigma^*, \circ)$ ist ein Monoid
- $(\mathbb{N} \mod 3, \oplus_3)$ ist eine Gruppe
- $(\triangle_D, \circ)$ ist eine Gruppe

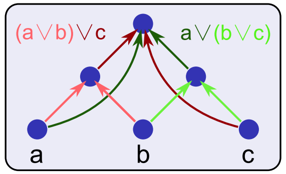


# Verbände

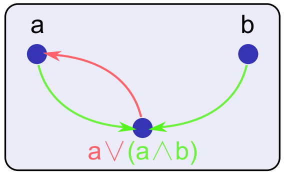
## Verband: algebraische Definition

Ein Verband $\mathbf{V} = (V, \vee, \wedge)$ ist eine Algebra, bestehend aus einer Menge $V$ und zwei binären Operationen $\vee$ und $\wedge$, die folgende Bedingungen erfüllen:

- Kommutativgesetze: $a \vee b = b \vee a$ und $a \wedge b = b \wedge a$
- Assoziativgesetze: $a \vee (b \vee c) = (a \vee b) \vee c$ und $a \wedge (b \wedge c) = (a \wedge b) \wedge c$
- Idempotenzgesetze: $a \vee a = a$ und $a \wedge a = a$
- Absorptionsgesetze: $a \vee (a \wedge b) = a$ und $a \wedge (a \vee b) = a$




$a \vee (b \vee c) = (a \vee b) \vee c$




$a \vee (a \wedge b) = a$



# Verbände

## Zusammenhang algebraischer und ordnungstheoretischer Verband

(i) $\mathbf{V} = (V, \trianglelefteq)$ sei ein (ordnungstheoretisch definierter) Verband. Setze $a \wedge b = \inf\{a, b\}$ und $a \vee b = \sup\{a, b\}$. Dann ist $\mathbf{V} = (V, \vee, \wedge)$ ein (algebraisch definierter) Verband.

(ii) $\mathbf{V} = (V, \vee, \wedge)$ sei ein (algebraisch definierter) Verband. Setze $a \trianglelefteq b$ g.d.w. $a \wedge b = a$. Dann ist $\mathbf{V} = (V, \trianglelefteq)$ ein (ordnungstheoretisch definierter) Verband.

Beispiel: $(\mathcal{POT}(M), \subseteq)$ und $(\mathcal{POT}(M), \cup, \cap)$



# Mathematische Grundlagen der Computerlinguistik

## Bäume

Dozentin: Wiebke Petersen

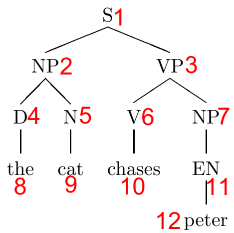
6. Foliensatz
(basierend auf Folien von Gerhard Jäger)

# Bäume

## Baumdiagramme

Ein Baumdiagramm eines Satzes stellt drei Arten von Information dar:

- die Konstituentenstruktur des Satzes,
- die grammatische Kategorie jeder Konstituente, sowie
- die lineare Anordnung der Konstituenten.






# Bäume

## Konventionen

- Ein Baum besteht aus Knoten, die durch
- Kanten verbunden werden.
- Kanten sind implizit von oben nach unten gerichtet (ähnlich zu Hasse-Diagrammen, wo die implizite Richtung aber von unten nach oben ist.)
- Jeder Knoten ist mit einem Etikett (engl. label) versehen.



# Bäume

## Dominanz

- Ein Knoten $x$ dominiert Knoten $y$ genau dann, wenn es eine zusammenhängende (möglicherweise leere) Sequenz von abwärts gerichteten Ästen gibt, die mit $x$ beginnt und mit $y$ endet.
- Für einen Baum $T$ bildet

$$D_T = \{\langle x, y\rangle | x \text{ dominiert } y \text{ in } T\}$$

  die zugehörige Dominanz-Relation.
- $D_T$ ist eine schwache Ordnung, also reflexiv, transitiv und anti-symmetrisch.

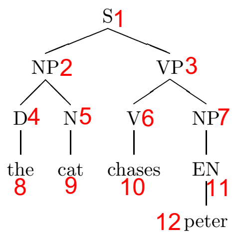


# Bäume

## Konventionen

- Wenn $x$ bzgl. $D_T$ der unmittelbare Vorgänger von $y$ ist, dann dominiert $x$ $y$ unmittelbar.
- Der unmittelbare Vorgänger von $x$ bzgl. $D_T$ heißt der Mutterknoten von $x$.
- Die unmittelbaren Nachfolger von $x$ heißen Tochterknoten von $x$.
- Wenn zwei Knoten nicht identisch sind, aber den selben Mutterknoten haben, heißen sie Schwesterknoten.
- Jeder Baum hat endlich viele Knoten.
- Jeder Baum hat ein Infimum bezüglich der Ordnung $D_T$. Das Infimum heißt Wurzel oder Wurzelknoten des Baums und dominiert alle anderen Knoten. Vorsicht: Die Baumdiagramme sind auf den Kopf gestellte Hasse-Diagramme (die Wurzel ist der oberste Knoten des Baumdiagramms, also der Knoten, der als einziges keinen Mutterknoten hat)
- Die maximalen Elemente eines Baumes heißen Blätter. Blätter stehen in einem Baumdiagramm ganz unten. Blätter sind diejenigen Knoten, die keine Töchter haben.

## Beispiel




- Knoten 2 dominiert Knoten 8 ($\langle 2, 8 \rangle \in D_T$)
- Knoten 2 dominiert Knoten 5 unmittelbar
- Knoten 2 dominiert Knoten 2
- Knoten 2 ist der Mutterknoten von Knoten 5
- Knoten 4 und Knoten 5 sind Schwesterknoten
- Knoten 1 ist der Wurzelknoten des Baums
- Knoten 10 ist ein Blatt des Baums



## Bäume

### Präzedenz

- Baum-Diagramme beinhalten (anders als Hasse-Diagramm) Informationen über die lineare Abfolge der Knoten.
- Knoten $x$ geht Knoten $y$ voran (engl. $x$ *precedes* $y$) genau dann, wenn $x$ links von $y$ steht und keiner der beiden Knoten den anderen dominiert.
- Für einen Baum $T$ bildet

$$P_T = \{\langle x, y \rangle | x \text{ geht } y \text{ voran}\}$$

  die zugehörige Präzedenz-Relation.
- $P_T$ ist eine starke Ordnung, also irreflexiv, transitiv und asymmetrisch.



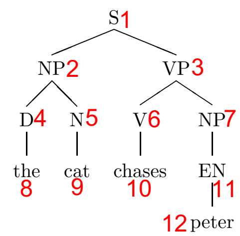
## Bäume

### Exklusivität

In einem Baum $T$ stehen die Knoten $x$ und $y$ in der Präzedenz-Relation (also $P_t(x, y)$ oder $P_t(y, x)$) genau dann, wenn sie nicht in der Dominanz-Relation stehen (also weder $D_T(x, y)$ noch $D_T(y, x)$).



## Beispiel




- Knoten 7 und Knoten 1 stehen in der Dominanz-Relation
- Knoten 7 und Knoten 2 stehen in der Präzedenz-Relation
- Knoten 7 und Knoten 9 stehen in der Präzedenz-Relation
- Knoten 7 und Knoten 12 stehen in der Dominanz-Relation
- Knoten 7 und Knoten 10 stehen in der Präzedenz-Relation

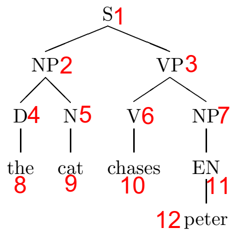
# Bäume

## Nicht-Überkreuzung

Wenn in einem Baum der Knoten $x$ dem Knoten $y$ vorangeht, dann geht jeder Knoten $x'$, der von $x$ dominiert wird, jedem Knoten $y'$ voran, der von $y$ dominiert wird.

Diese Bedingung schließt folgende Situationen aus:

- Ein Knoten hat mehrere Mutterknoten.
- Äste überkreuzen sich.



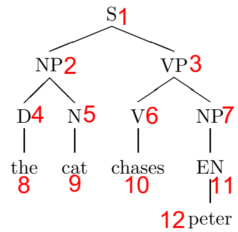



# Bäume

## Etikettierung

Eine Etikettierungsfunktion $L_T$ eines Baums $T$ ist eine Funktion, die jedem Knoten ein Etikett zuweist.

- $L_T$ muss nicht injektiv sein (mehrere Knoten können das selbe Etikett tragen).
- Bei Ableitungsbäumen werden Blätter (auch Terminal-Knoten genannt) auf Terminalsymbole abgebildet und alle anderen Knoten auf Nichtterminal-Symbole.






# Bäume

Mit Hilfe dieser Eigenschaften von Bäumen können *Theoreme* bewiesen werden, also Sachverhalte, die für alle Bäume gelten. Zum Beispiel

## Theorem

*Wenn $x$ und $y$ Schwesterknoten sind, dann gilt entweder $P_T(x, y)$ oder $P_T(y, x)$.*

Beweisskizze: Schwesterknoten haben dieselbe Mutter und stehen untereinander nicht in der Dominanzrelation. Aus dem Exklusivitätssatz folgt, dass sie in der Präzedenzrelation stehen.

## Theorem

*Die Menge der Blätter eines Baumes sind durch $P_T$ total geordnet.*

Beweisskizze: Folgt aus dem Satz der Nicht-Überkreuzung.



# Grammatiken und Bäume

- Zur Erinnerung: kontextfreie Grammatiken sind Grammatiken mit Regeln, deren linke Regelseiten aus genau einem Nichtterminalsymbol bestehen:

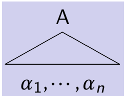
$$A \to \alpha$$

  mit $A \in N$ und $\alpha \in (T \cup N)^*$
- Ableitungen kontextfreier Grammatiken können als etikettierte Bäume abgebildet werden.
- Bäume repräsentieren dabei die relevanten Aspekte einer Ableitung (also welche Regeln für die Generierung welcher Konstituenten angewandt wurden, aber nicht, in welcher Reihenfolge Regeln angewandt wurden).

## Grammatiken und Bäume

### Definition

Eine kontextfreie Grammatik $G = (N, T, S, P)$, bei der alle Regeln als linke Seite genau ein Nichtterminalsymbol haben, generiert einen Baum $B$ genau dann, wenn

- die Wurzel von $B$ mit $S$ etikettiert ist,
- die Blätter entweder mit Terminalsymbolen oder mit $\epsilon$ etikettiert sind, sowie
- es für jeden Teilbaum 

in $B$ eine Regel $A \to \alpha_1, \cdots, \alpha_n$ in $P$ gibt.

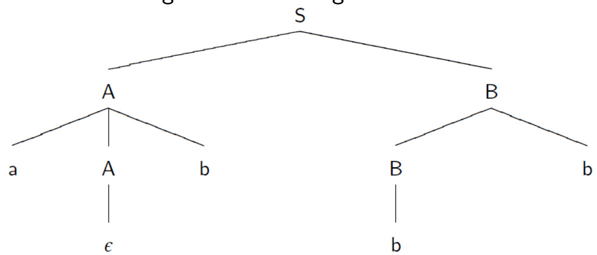


## Grammatiken und Bäume

### Beispielgrammatik

$$G = (\{S, A, B\}, \{a, b\}, S, P) \quad P = \left\{ \begin{array}{ll} S \to AB & B \to Bb \\ A \to aAb & B \to b \\ A \to \epsilon & \end{array} \right\}$$

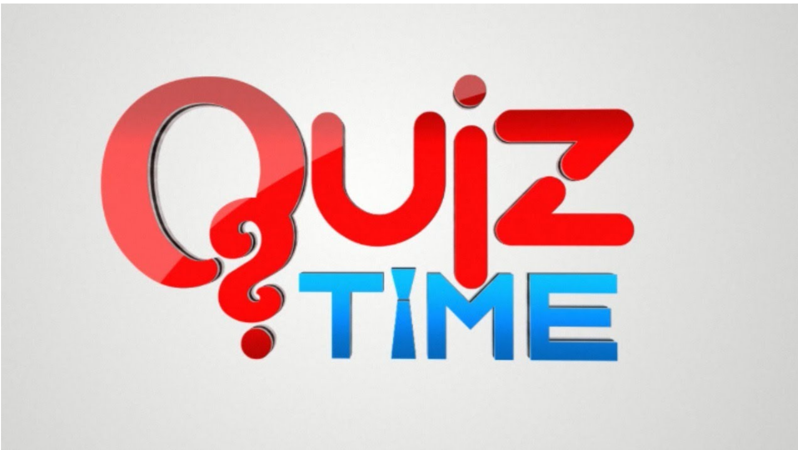
Diese Grammatik generiert z.B. folgenden Baum:




Frage: Welche Sprache wird durch diese Grammatik generiert?



## Quiz-Time






# Mathematische Grundlagen der Computerlinguistik

## Kombinatorik

Dozentin: Wiebke Petersen

7. Foliensatz

# Kombinatorik

- Thema der Kombinatorik ist die Bestimmung der Anzahl möglicher Anordnungen oder Auswahlen.
- Typische kombinatorische Aufgaben sind Urnenaufgaben: Wieviele Möglichkeiten gibt es $k$ Kugeln aus einer Urne mit $n$ Kugeln zu ziehen?
- Hierbei unterscheidet man
  - ob die gezogenen Kugeln wieder zurückgelegt werden oder nicht, und
  - ob die Reihenfolge, in der die Kugeln gezogen werden, beachtet wird oder nicht.



# kombinatorische Grundaufgaben: Beispiele

| | ohne Zurücklegen | mit Zurücklegen |
|---|---|---|
| mit Beachtung der Reihenfolge | 3er-Wette (Rennsport) | Toto |
| ohne Beachtung der Reihenfolge | Lotto / Skat | Eisbecher |



# Ziehen mit Zurücklegen mit Beachtung der Reihenfolge

Beispiel: Tippen der Ergebnisse von 11 Fußballspielen (1: Sieg Heimmannschaft, 2: Sieg Gastmannschaft, 0: unentschieden).

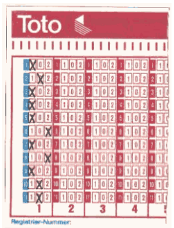
$$3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 3^{11} = 177147$$

Es gibt

$$n^k$$

Möglichkeiten $k$ Objekte aus einer Menge von $n$ Objekten mit Beachtung ihrer Reihenfolge und mit Zurücklegen auszuwählen.

**Beispiel: Toto (11er-Wette)**






# Ziehen ohne Zurücklegen mit Beachtung der Reihenfolge

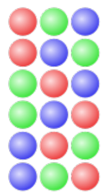
Spezialfall: alle Kugeln werden gezogen ($n = k$)

## Permutationen

$n$ Objekte lassen sich auf $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$ verschiedene Arten in einer Reihe anordnen.
Der Ausdruck $n!$ wird ‚$n$ Fakultät'gelesen.
Als Permutation bezeichnet man eine bijektive Abbildung einer endlichen Menge auf sich selbst.
Zu einer $n$-elementigen Menge gibt es $n!$ Permutationen.
Permutationen sind ein Spezialfall ($k = n$) des ‚Ziehens ohne Zurücklegen mit Beachtung der Reihenfolge'

Lineare Anordnungsmöglichkeiten für 3 verschiedenfarbige Kugeln:

$$3! = 1 \cdot 2 \cdot 3 = 6$$



## Ziehen ohne Zurücklegen mit Beachtung der Reihenfolge

Beispiel: Tippen der ersten 3 Plätze bei einem Pferderennen, wenn 10 Pferde starten.

$$10 \cdot 9 \cdot 8 = \frac{10!}{(10-3)!} = 720$$

Es gibt $\frac{n!}{(n-k)!}$

Möglichkeiten $k$ Objekte aus einer Menge von $n$ Objekten mit Beachtung ihrer Reihenfolge und ohne Zurücklegen auszuwählen.

**Beispiel: 3er-Wette Pferderennsport**



## Ziehen ohne Zurücklegen ohne Beachtung der Reihenfolge

Beispiel: Lottospiel (6 aus 49)

$$\frac{49!}{(49-6)!} : 6! = \frac{49!}{(49-6)! \cdot 6!} = \binom{49}{6} = 13983816$$

Beispiel: Skathände (10 aus 32)

$$\frac{32!}{(32-10)!} : 10! = \frac{32!}{(32-10)! \cdot 10!} = \binom{32}{10} = 64512240$$

Es gibt $\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$

Möglichkeiten $k$ Objekte aus einer Menge von $n$ Objekten ohne Beachtung ihrer Reihenfolge und ohne Zurücklegen auszuwählen.

**Beispiel: Lotto**

**Beispiel: Skat**



## Herleitung: ohne Reihenfolge, ohne Zurücklegen

Es gibt $\frac{n!}{(n-k)!}$

Möglichkeiten $k$ Objekte aus einer Menge von $n$ Objekten ohne Zurücklegen aber mit Beachtung der Reihenfolge auszuwählen.

Jede $k$-Auswahl ohne Wiederholungen lässt sich auf $k!$ Arten anordnen.

Folglich gibt es

$$\frac{n!}{(n-k)!} : k! = \frac{n!}{(n-k)!} \cdot \frac{1}{k!} = \frac{n!}{k! \cdot (n-k)!}$$

verschiedene ungeordnete $k$-Auswahlen aus einer $n$er-Menge ohne Wiederholungen.

Die Zahlen $\frac{n!}{k! \cdot (n-k)!}$ sind die Binomialkoeffizienten und werden oft mit $\binom{n}{k}$ bezeichnet (in Worten ‚$n$ über $k$').

Es gibt $\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$

Möglichkeiten $k$ Objekte aus einer Menge von $n$ Objekten ohne Beachtung ihrer Reihenfolge und ohne Zurücklegen auszuwählen.



## Ziehen mit Zurücklegen ohne Beachtung der Reihenfolge

Beispiel: Eisbecher mit 3 Kugeln aus 10 Eissorten zusammenstellen.

$$\binom{10+3-1}{3} = 220$$

Es gibt $\binom{n+k-1}{k}$

Möglichkeiten $k$ Objekte aus einer Menge von $n$ Objekten ohne Beachtung ihrer Reihenfolge und mit Zurücklegen auszuwählen.

**Beispiel: Eisbecher**

## Herleitung: ohne Reihenfolge, mit Zurücklegen

Beispiel: gemischte Eisbecher mit 3 Kugeln aus 5 Eissorten

| Schoko | Nuß | Orange | Erdbeer | Banane | Becher | Kodierung |
|---|---|---|---|---|---|---|
| | • • | | | • | ●●● | \| • •\|\|\|• |
| | | • | • | • | ●●● | \|\|•\|•\|• |
| • | | | • • | | ●●● | •\|\|\|••\| |
| | | | • • • | | ●●● | \|\|\|•••\| |

Die Kodierung der Eisbecher ist so gewählt, dass sich das Problem der Wahl von $k$ Eiskugeln aus $n$ Eissorten auf das Problem der linearen Anordnung von $k$ ununterscheidbaren Kugeln und $n-1$ ununterscheidbaren Strichen reduziert. Dieses Problem lässt sich als Auswahl von $k$ Positionen (die Kugelpositionen) aus $k+n-1$ Positionen auffassen.

Hierfür gibt es $\binom{k+n-1}{k}$ Möglichkeiten

.



## kombinatorische Grundaufgaben:Zusammenfassung

Anzahl der $k$-Auswahlen aus einer $n$er-Menge:

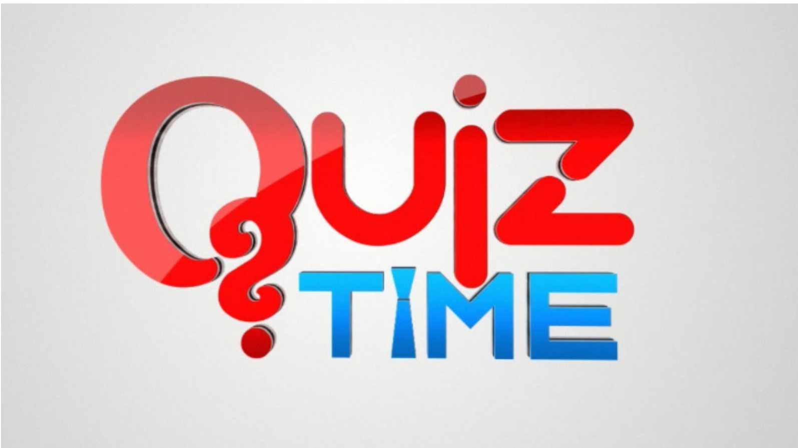
| | ohne Wiederholungen | mit Wiederholungen |
|---|---|---|
| mit Beachtung der Reihenfolge | $\binom{n}{k} \cdot k! = \frac{n!}{(n-k)!}$ | $n^k$ |
| ohne Beachtung der Reihenfolge | $\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$ | $\binom{n+k-1}{k}$ |

**Hinweise:**

- Bearbeiten Sie bitte das Modul Kombinatorik (Link)
- Berechnung von Binomialkoeffizienten (Link)



## Quiz-Time






# Mathematische Grundlagen der Computerlinguistik

## Wahrscheinlichkeit

Dozentin: Wiebke Petersen

8. Foliensatz

## Motivation

- In vielen Bereichen der CL kommt Wahrscheinlichkeitstheorie zur Anwendung, da es oft unmöglich ist, mit rein symbolischen Ansätzen ein vollständiges Bild aller möglichen Strukturen einschließlich Präferenzen bei Ambiguitäten zu gewinnen.
- Wir haben es meist mit einer endlichen oder abzählbar unendlichen Menge von sogenannten Ergebnissen zu tun, deren Wahrscheinlichkeit irgendwie abgeschätzt werden muss.

Bsp.:

- Wahrscheinlichkeit dafür, dass VP $\rightarrow$ VP PP verwendet wird, vorausgesetzt, man möchte eine VP generieren.
- Wahrscheinlichkeit dafür, dass *chair* eine Nomen ist.



## ideales Zufallsexperiment (Modell)

### Anforderungen an ein ideales Zufallsexperiment:

- Das Experiment wird unter genau festgelegten Versuchsbedingungen durchgeführt.
- Die Menge der möglichen Ergebnisse ist vor der Durchführung des Experiments bekannt.
- Das Experiment kann zumindest prinzipiell beliebig oft unter gleichen Bedingungen wiederholt werden.

### Ergebnisraum

Die Menge der möglichen Ergebnisse eines idealen Zufallsexperiments bildet den Ergebnisraum und wird mit $\Omega$ ('Omega') bezeichnet.
$\Omega$ wird auch der Stichprobenraum genannt.
Ist der Ergebnisraum nicht leer und abzählbar, dann heißt er diskret.



## Zufallsexperiment und Ereignisse

Wir unterscheiden einzelne Ergebnisse und Ereignisse, die Mengen von Ergebnissen sind.

- Ein Ereignis bildet eine Teilmenge von $\Omega$.
- $\emptyset$ ist das unmögliche Ereignis.
- $\Omega$ ist das sichere Ereignis.
- Zwei Ereignisse $E_1$ und $E_2$ heißen unvereinbar, wenn $E_1 \cap E_2 = \emptyset$.
- Die Einermengen $\{e\}$ $(e \in \Omega)$ heißen Elementarereignisse.
- Das Komplement eines Ereignisses $E$, also $\overline{E}$, heißt Gegenereignis zu $E$.

**Beispiel Zufallsexperiment: Würfeln mit einem Würfel**

- $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$
- Der Wurf einer 3 ist das Elementarereignis {⚂} des Zufallexperiments.
- {⚁, ⚃, ⚅} ist das Ereignis ‚Wurf einer geraden Augenzahl'
- Das Gegenereignis von ‚Wurf einer geraden Augenzahl'ist ‚Wurf einer ungeraden Augenzahl'



## Beispiel Zufallexperiment: Augensumme bei zweimaligem Würfeln

| | | |
|---|---|---|
| Summe 2 | entspricht dem Ereignis | {⚀⚀} |
| Summe 3 | entspricht dem Ereignis | {⚀⚁, ⚁⚀} |
| Summe 4 | entspricht dem Ereignis | {⚀⚂, ⚂⚀, ⚁⚁} |
| Summe 5 | entspricht dem Ereignis | {⚀⚃, ⚃⚀, ⚁⚂, ⚂⚁} |
| Summe 6 | entspricht dem Ereignis | {⚀⚄, ⚄⚀, ⚁⚃, ⚃⚁, ⚂⚂} |
| Summe 7 | entspricht dem Ereignis | {⚀⚅, ⚅⚀, ⚁⚄, ⚄⚁, ⚂⚃, ⚃⚂} |
| Summe 8 | entspricht dem Ereignis | {⚁⚅, ⚅⚁, ⚂⚄, ⚄⚂, ⚃⚃} |
| Summe 9 | entspricht dem Ereignis | {⚂⚅, ⚅⚂, ⚃⚄, ⚄⚃} |
| Summe 10 | entspricht dem Ereignis | {⚃⚅, ⚅⚃, ⚄⚄} |
| Summe 11 | entspricht dem Ereignis | {⚄⚅, ⚅⚄} |
| Summe 12 | entspricht dem Ereignis | {⚅⚅} |

## Diskrete Wahrscheinlichkeitsräume

Ein diskreter Wahrscheinlichkeitsraum ist ein Paar $\langle \Omega, P \rangle$, bestehend aus

1. einer nicht leeren, abzählbaren Menge $\Omega$ von Ergebnissen (diskreter Ergebnisraum) und
2. einem Wahrscheinlichkeitsmaß $P : \mathcal{POT}(\Omega) \to \mathbb{R}$, so dass
   1. $P(A) \geq 0$ für alle $A \in \mathcal{POT}(\Omega)$;
   2. $P(\Omega) = 1$;
   3. für paarweise disjunkte Mengen $A_n \in \mathcal{POT}(\Omega)$, $n \in \mathbb{N}$ gilt

$$P(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$$



## Diskrete Wahrscheinlichkeitsräume

Es ergeben sich folgende Eigenschaften für Wahrscheinlichkeitsmaße:

1. $P(\emptyset) = 0$
2. Für Ereignisse $A$, $B$ mit $A \cap B = \emptyset$ gilt $P(A \cup B) = P(A) + P(B)$.
3. $P(A) + P(\overline{A}) = 1$ für alle $A \subseteq \Omega$ (Tertium non datur)
4. Impliziert Ereignis $A$ das Ereignis $B$ (d.h. $A \subseteq B$), dann gilt $P(B \setminus A) = P(B) - P(A)$.
5. Kein Ereignis kann eine Wahrscheinlichkeit über 1 haben.



## Diskrete Wahrscheinlichkeitsräume

Bsp.: $\Omega = \{$is-noun, has-plural-s, is-adjective, is-verb$\}$.
Frage: Kann die Funktion $f$ mit

$$\begin{aligned} f(\text{is-noun}) &= 0.45 \\ f(\text{has-plural-s}) &= 0.2 \\ f(\text{is-adjective}) &= 0.25 \\ f(\text{is-verb}) &= 0.3 \end{aligned}$$

zu einem Wahrscheinlichkeitsmaß $f : \mathcal{P}(\Omega) \to \mathbb{R}$ ergänzt werden?
Nein, da dann $f(\Omega) = 0.45 + 0.2 + 0.25 + 0.3 = 1.2 > 1$ wäre.



## Diskrete Wahrscheinlichkeitsräume

Besser: $\Omega = \{$is-noun-with-plural-s, is-noun-without-plural-s, is-adjective, is-verb$\}$.

$$\begin{aligned} f(\text{is-noun-with-plural-s}) &= 0.09 \\ f(\text{is-noun-without-plural-s}) &= 0.36 \\ f(\text{is-adjective}) &= 0.25 \\ f(\text{is-verb}) &= 0.3 \end{aligned}$$

# Laplace-Räume

Laplace-Räume sind diskrete Wahrscheinlichkeitsräume, in denen alle Ergebnisse gleich wahrscheinlich sind.
Bsp.: Würfelexperiment. $\Omega = \{1, 2, 3, 4, 5, 6\}$.
Jedes Ergebnis hat die Wahrscheinlichkeit $\frac{1}{|\Omega|} = \frac{1}{6}$.

In Laplace-Räumen gilt also

$$P(A) = \frac{|A|}{|\Omega|}$$

.



# Beispiel Laplace-Raum: zweimaliges Würfeln und Augensumme

| Augensumme | Ereignis | Wahrscheinlichkeit |
|---|---|---|
| 2 | {⚀⚀} | $\frac{1}{36}$ |
| 3 | {⚀⚁, ⚁⚀} | $\frac{2}{36}$ |
| 4 | {⚀⚂, ⚂⚀, ⚁⚁} | $\frac{3}{36}$ |
| 5 | {⚀⚃, ⚃⚀, ⚁⚂, ⚂⚁} | $\frac{4}{36}$ |
| 6 | {⚀⚄, ⚄⚀, ⚁⚃, ⚃⚁, ⚂⚂} | $\frac{5}{36}$ |
| 7 | {⚀⚅, ⚅⚀, ⚁⚄, ⚄⚁, ⚂⚃, ⚃⚂} | $\frac{6}{36}$ |
| 8 | {⚁⚅, ⚅⚁, ⚂⚄, ⚄⚂, ⚃⚃} | $\frac{5}{36}$ |
| 9 | {⚂⚅, ⚅⚂, ⚃⚄, ⚄⚃} | $\frac{4}{36}$ |
| 10 | {⚃⚅, ⚅⚃, ⚄⚄} | $\frac{3}{36}$ |
| 11 | {⚄⚅, ⚅⚄} | $\frac{2}{36}$ |
| 12 | {⚅⚅} | $\frac{1}{36}$ |



# Beispiel Laplace-Raum: Geburtstagsproblem

Bsp.: Wahrscheinlichkeit dafür, dass in einer Gruppe von 30 Personen mindestens zwei am gleichen Tag Geburtstag haben.
Vereinfachung: Wir ignorieren Schaltjahre und saisonale Variationen.
D.h., Wahrscheinlichkeit dafür, an einem bestimmten Tag Geburtstag zu haben, ist $\frac{1}{365}$.
Wahrscheinlichkeitsraum:

- $\Omega = \{1, \ldots, 365\}^{30}$, also alle Folgen von 30 Zahlen aus $\{1, \ldots, 365\}$.
- $|\Omega| = 365^{30}$. Alle Folgen sind gleichwahrscheinlich (Laplace-Raum).



# Hinweis

Für die Modellierung als Laplace-Raum ist es unerläßlich, die Geburtstagsverteilung als Urnenproblem mit Beachtung der Reihenfolge zu betrachten.
Würde die Reihenfolge vernachlässigt und $\Omega$ als die Menge aller ungeordneten Kombinationen möglicher Geburtstagsverteilungen betrachtet (also $|\Omega| = \binom{n+k-1}{k}$), so wären die Ergebnisse in dem Ergebnisraum nicht gleichwahrscheinlich.
Beispiel:
Die Wahrscheinlichkeit für das Ergebnis, dass alle am 1. Januar Geburtstag haben ist $\left(\frac{1}{365}\right)^{30}$ während die Wahrscheinlichkeit dafür, dass die Geburtstage genau auf die ersten 30 Tage des Jahres fallen $\left(\frac{1}{365}\right)^{30} * 30!$ ist.

## Beispiel Laplace-Raum: Geburtstagsproblem

Ziel: Wahrscheinlichkeit dafür, dass eine Folge eintritt, in der sich mindestens ein Element wiederholt.
Einfacher: Wahrscheinlichkeitsermittlung über das Komplement.
Wieviel Folgen gibt es, in denen sich kein Element wiederholt?

$$365 \times 364 \times \cdots \times (365 - 29) = \frac{365!}{(365-30)!}$$

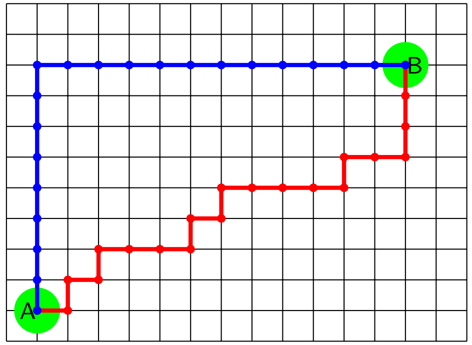
$\Rightarrow$ Wahrscheinlichkeit dafür, dass zwei am gleichen Tag Geburtstag haben ist

$$1 - \frac{365!}{365^{30}(365-30)!} \approx 1 - 0.29 = 0.71$$



## Beispiel Laplace-Raum: Wege im Gitter

Wege von $A$ nach $B$. Regel: Schritt nach oben oder Schritt nach rechts.
Beide Richtungen gleichwahrscheinlich.

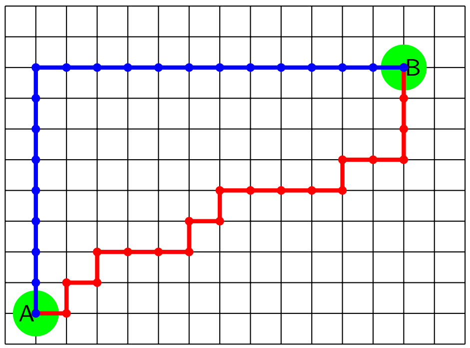



Wieviele Wege gibt es von $A$ nach $B$?
Jeder erfolgreiche Weg besteht aus 20 Schritten (8 nach oben, 12 nach rechts).
Zahl der möglichen Wege: $|A \rhd^{20} B| = \binom{20}{8} = 125970$ (Bild deckt nicht alle Möglichkeiten ab)



## Beispiel Laplace-Raum: Wege im Gitter

Wege von $A$ nach $B$. Regel: Schritt nach oben oder Schritt nach rechts.
Beide Richtungen gleichwahrscheinlich.

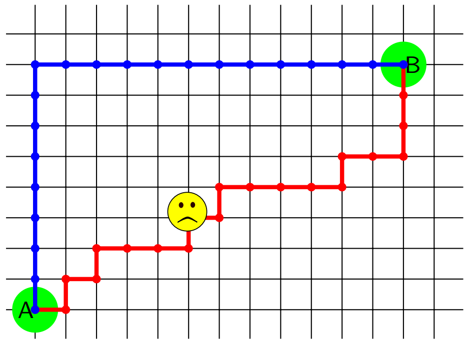



Wie hoch ist die Wahrscheinlichkeit für das Ereignis nach 20 Schritten bei Punkt $B$ anzukommen?

$$P(A \rhd^{20} B) = \frac{|A \rhd^{20} B|}{|\Omega|} = \frac{\binom{20}{8}}{2^{20}} = \frac{125970}{1048576} \approx 0,12$$

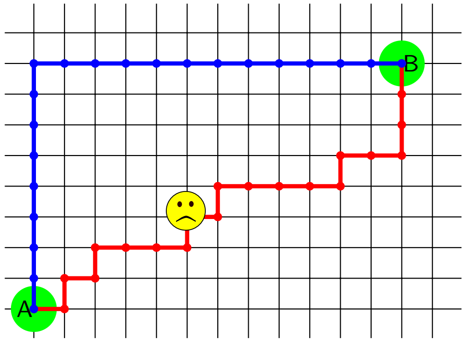


## Beispiel Laplace-Raum: Gitter mit Hindernis




Wie hoch ist die Wahrscheinlichkeit für das Ereignis auf einem Weg von $A$ nach $B$ nicht auf das Hindernis zu stoßen? Also die Wahrscheinlichkeit unter allen Wegen von $A$ nach $B$ einen zu wählen, der nicht auf das Hindernis stößt.
Wege von A nach B, die das Hindernis treffen: $\binom{8}{3} \cdot \binom{12}{5} = 56 \cdot 792 = 44352$
Wahrscheinlichkeit Weg von A nach B ohne Hindernis: $\frac{125970-44352}{125970} \approx 0,65$

## Beispiel Laplace-Raum: Gitter mit Hindernis

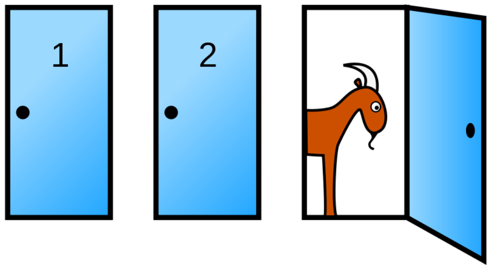



Wie hoch ist die Wahrscheinlichkeit für das Ereignis nach 20 Schritten sowohl bei Punkt $B$ anzukommen, als auch auf dem Weg nicht auf das Hindernis zu treffen?

$$\frac{\binom{20}{8} - \binom{8}{3} \cdot \binom{12}{5}}{2^{20}} = \frac{125970 - 44352}{1048576} \approx 0,078$$



## Ziegenproblem




Situation: 3 verschlossene Türen, hinter einer der Türen befindet sich ein Gewinn, hinter zwei Türen befinden sich Nieten (Ziegen).

1. der Kandidat wählt eine Tür
2. der Moderator öffnet von den verbleibenden beiden Türen eine Ziegentür
3. der Kandidat darf die Tür wechseln

Frage: Lohnt sich ein Wechsel?



## Bedingte Wahrscheinlichkeiten

Bsp.:

- Wahrscheinlichkeit für eine Produktion VP $\rightarrow$ V NP für die Generierung einer VP, gegeben, dass es sich um das Verb *kisses* (bzw. *sleeps*) handelt.
- Wahrscheinlichkeit dafür, dass *chairs* ein Nomen ist, gegeben die Tatsache, dass das vorangehende Wort ein Artikel ist.
- Wahrscheinlichkeit dafür, dass *chairs* ein Nomen ist, gegeben die Tatsache, dass das nachfolgende Wort ein Artikel ist.



## **Bedingte Wahrscheinlichkeiten**

In einem diskreten Wahrscheinlichkeitsraum $\langle \Omega, P \rangle$, gegeben ein Ereignis $B \subseteq \Omega$ mit $P(B) > 0$, ist durch

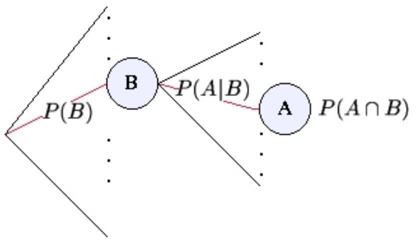
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

das durch $B$ bedingte Wahrscheinlichkeitsmaß $P(\cdot|B) : \Omega \rightarrow \mathbb{R}$ auf $\mathcal{P}(\Omega)$ definiert.

$\langle \mathcal{P}(\Omega), P(\cdot|B) \rangle$ ist ein diskreter Wahrscheinlichkeitsraum.

$P(A|B)$ wird gelesen als 'die Wahrscheinlichkeit von A gegeben B' und steht für die Wahrscheinlichkeit, dass das Ereignis A eintritt, gegeben die Sicherheit, dass das Ereignis B eingetreten ist.

## bedingte Wahrscheinlichkeiten: Produktregel




**Produktregel**

$$P(A \cap B) = P(B) \cdot P(A|B)$$



## Unabhängigkeit von Ereignissen

Zwei Ereignisse $A$ und $B$ heißen unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$. Das heißt $P(A|B) = P(A)$.

Bsp. Würfelexperiment.

- Die Ereignisse, dass ($A$) eine gerade Zahl gewürfelt wird und ($B$) eine Zahl $\leq 2$, sind unabhängig:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{1,2\})} = 0.5 = P(A)$$

- Die Ereignisse $A$ wie oben und $B$, dass genau die 2 gewürfelt wird, sind nicht unabhängig:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{2\})} = 1 \neq P(A)$$



## Die Formel von Bayes

Thomas Bayes (1701-1761)



## Die Formel von Bayes

Ziel: $P(A|B)$ berechnen auf der Grundlage von $P(B|A)$, $P(A)$ und $P(B)$.

Laut Definition gilt

$$P(A \cap B) = P(A|B) \cdot P(B) \text{ und } P(B \cap A) = P(B|A) \cdot (P(A)$$

Daraus ergibt sich

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

# Die Formel von Bayes

Man kann das Theorem von Bayes noch verallgemeinern: Angenommen, es gibt eine endliche oder abzählbar unendliche Folge $(A_i)_{i\in\mathbb{N}}$ von paarweise disjunkten Ereignissen mit $A_i \subseteq \Omega$ und $P(A_i) > 0$ für alle $i \in \mathbb{N}$, die eine Zerlegung von $\Omega$ bilden, dann gilt für jedes Ereignis $B \subseteq \Omega$: $(B \cap A_i)_{i\in\mathbb{N}}$ bildet eine disjunkte Zerlegung von $B$, und daher

$$P(B) = \sum_{i\in\mathbb{N}} P(B|A_i)P(A_i)$$

Spezialfall: Zerlegung in $A$ und $\overline{A}$:

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$$

$$P(B) = \frac{P(B\cap A)}{P(A)}P(A) + \frac{P(B\cap \overline{A})}{P(\overline{A})}P(\overline{A})$$

$$P(B) = P(B\cap A) + P(B\cap \overline{A})$$



# Die Formel von Bayes

Aus

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

und

$$P(B) = \sum_{i\in\mathbb{N}} P(B|A_i)P(A_i)$$

ergibt sich dann für die Folge $(A_i)_{i\in\mathbb{N}}$ und das Ereignis $B$ wie oben die verallgemeinerte Formel von Bayes:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i\in\mathbb{N}} P(B|A_i)P(A_i)}$$



# Die Formel von Bayes

Bsp.: Angenommen, wir interessieren uns für eine relativ seltene Konstruktion, z.B. *Parasitic Gaps*, die ungefähr alle 100.000 Sätze einmal vorkommt.[1] Joe Linguist hat einen Pattern-Matching Algorithmus zur Erkennung von Parasitic Gaps implementiert, der, falls ein Satz ein Parasitic Gap enthält, dies mit einer Wahrscheinlichkeit von 0.95 auch erkennt. Enthält ein Satz kein Parasitic Gap, liefert der Algorithmus mit einer Wahrscheinlichkeit von 0.005 das falsche Ergebnis, dass ein Parasitic Gap in dem Satz vorhanden ist.
Frage: Angenommen, der Test meldet ein Parasitic Gap in einem Satz. Wie wahrscheinlich ist es, dass es sich wirklich um eines handelt?

[1] Z.B. *which book did she review _ without reading _?*.



# Die Formel von Bayes

Sei $G$ das Ereignis eines parasitic gaps, $T$ das eines positiven Tests.
Wir kennen die Werte $P(G) = \frac{1}{100.000} = 0,00001$,
$P(\overline{G}) = 1 - P(G) = 0,99999$, $P(T|G) = 0,95$ und $P(T|\overline{G}) = 0,005$.
Wir wollen $P(G|T)$ berechnen.

$$P(G|T) = \frac{P(G\cap T)}{P(T)} = \frac{P(T|G)\cdot P(G)}{P(T)}$$

$P(T)$ lässt sich über $P(T|G)$ und $P(T|\overline{G})$ berechnen:

$$P(T) = P(T\cap G) + P(T\cap \overline{G}) = P(T|G)\cdot P(G) + P(T|\overline{G})\cdot P(\overline{G})$$

Also erhalten wir

$$P(G|T) = \frac{P(T|G)P(G)}{P(T|G)P(G) + P(T|\overline{G})P(\overline{G})}$$

$$= \frac{0.95 \times 0.00001}{0.95 \times 0.00001 + 0.005 \times 0.99999}$$

$$\approx 0.002$$

## Gefangenenparadoxon

Aus drei zum Tode verurteilten Gefangene (Anton, Bernd und Clemens) wird einer zur Begnadigung ausgewählt. Anton erfährt, dass Bernd hingerichtet wird. Anton erzählt dies Clemens weiter.

- Anton: entweder Clemens wird begnadigt oder er selbst, so dass seine Überlebenschance von $1/3$ auf $1/2$ gestiegen sei.
- Clemens: Überlebenschance von $1/3$ auf $2/3$ gestiegen

Wer von beiden Gefangenen schätzt seine Chancen korrekt ein?



## Berkeley 1973 (Simpsonsche Paradoxon)

- Annahmequote Universität
  für Männer: $44,3\%$ für Frauen: $34,6\%$
- liegt hier eine Benachteiligung der Frauen vor?

Nicht notwendig, siehe: Statistikmodul Mathe-Prisma (Link)

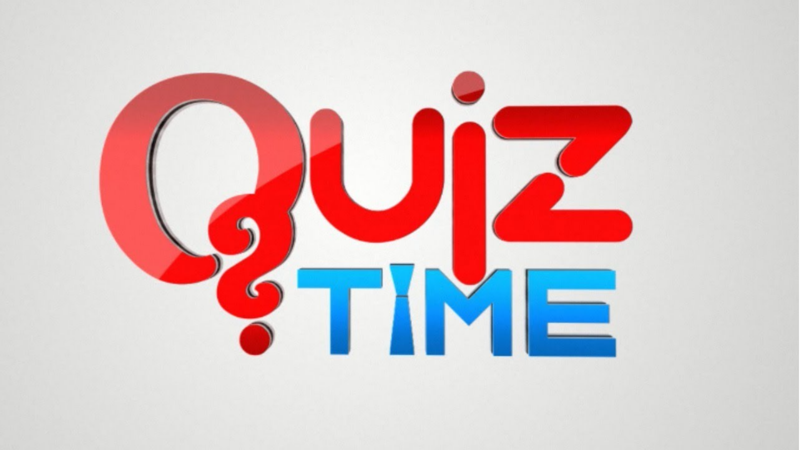


## Hinweis

Arbeiten Sie bitte das Mathe-Prisma Modul zur bedingten Wahrscheinlichkeit durch (Link)



## Quiz-Time



## Mathematische Grundlagen der Computerlinguistik

## Lineare Algebra

Dozentin: Wiebke Petersen

9. Foliensatz



## Einleitung

Die lineare Algebra beschäftigt sich mit

- Vektorräumen
- linearen Abbildungen zwischen Vektorräumen
- linearen Gleichungssystemen
- Matrizen



## Gruppen (Wiederholung)

Eine Gruppe $\mathbf{G} = (G, \otimes)$ ist eine algebraische Struktur mit:

G1 $\otimes$ ist assoziativ

G2 $G$ enthält ein neutrales Element

G3 jedes Element aus $G$ hat ein inverses Element in $G$.

**G** heißt abelsche Gruppe, wenn die Verknüpfung $\otimes$ kommutativ ist.



## Körper

Ein Körper $\mathbf{K} = (K, +, \cdot)$ ist eine algebraische Struktur mit zwei Verknüpfungen $+ : K \times K \to K$ (Addition) und $\cdot : K \times K \to K$ (Multiplikation):

K1 $(K, +)$ ist eine abelsche Gruppe.
Das neutrale Element der Addition wird mit 0 bezeichnet und das zu $a$ inverse Element mit $-a$.

K2 Die Multiplikation lässt sich auf $K \setminus \{0\}$ beschränken (für $a, b \in K \setminus \{0\}$ gilt $a \cdot b \in K \setminus \{0\}$), und $K \setminus \{0\}$ zusammen mit der Multiplikation bildet eine abelsche Gruppe.
Das neutrale Element der Multiplikation wird mit 1 bezeichnet und das zu $a \in K \setminus \{0\}$ inverse Element mit $a^{-1}$.

K3 Es gelten die folgenden Distributivgesetze:
$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ und $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$.

## Körper: Anmerkungen und Beispiele

- Statt $a \cdot b$ schreibt man zumeist $ab$.
- Per Konvention bindet die Multiplikation stärker als die Addition, es gilt also $(ab) + c = ab + c$.
- Statt $a + (-b)$ schreibt man zumeist $a - b$.
- Es gelten folgende Zusammenhänge (versuchen sie die Aussagen zu beweisen):
  - $0 \cdot a = a \cdot 0 = 0$ für alle $a \in K$.
  - Ist $ab = 0$ so gilt $a = 0$ oder $b = 0$.
  - Für alle $a, b \in K$ gilt $a(-b) = (-a)b = -(ab)$. Außerdem gilt $(-a)(-b) = ab$.

Beispiele:

- $(\mathbf{Q}, +, \cdot)$ ist ein Körper.
- $(\mathbb{R}, +, \cdot)$ ist ein Körper.
- Aber $(\mathbf{Z}, +, \cdot)$ ist kein Körper.



## Beweise einiger Eigenschaften von Körpern

Sei $(K, +, \cdot)$ ein Körper.

- Aussage: $0 \cdot a = a \cdot 0 = 0$ für alle $a \in K$
  Beweis: $0 \cdot a = (0 + 0) \cdot a$ (neutrales Element der Addition)
  $= 0 \cdot a + 0 \cdot a$ (Distributivgesetz)
  Also: $0 \cdot a = 0 \cdot a + 0 \cdot a$. Daraus folgt, dass $0 \cdot a$ das neutrale Element der Addition ist, also 0. Der Beweis, dass $a \cdot 0 = 0$ erfolgt analog.$\Box$
- Aussage: Ist $ab = 0$ so gilt $a = 0$ oder $b = 0$.
  Beweis durch Fallunterscheidung: 1. Fall: Wenn $a = 0$ und $a \cdot b = 0$ dann folgt die Aussage direkt. 2. Fall: Wenn $a \neq 0$ und $a \cdot b = 0$, dann gilt $b = 1 \cdot b = (a^{-1} \cdot a) \cdot b = a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0 = 0$. $\Box$
- Aussage: Für alle $a, b \in K$ gilt $a(-b) = (-a)b = -(ab)$
  Beweis: $ab + (-a)b = (a - a)b = 0b = 0$ ($(-a)b$ ist inverses Element von $ab$)
  $ab + a(-b) = a(b - b) = a \cdot 0 = 0$. Daraus folgt
  $a(-b) = (-a)b = -(ab)$.$\Box$
- Aussage: $(-a)(-b) = ab$
  Beweis:
  $0 = -ab - (-ab) = -ab - (-a)b = -ab + (-a)(-b) = (-a)(-b) - ab = 0$
  ($-ab$ ist das inverse Element von $(-a)(-b)$).$\Box$



## Vektorraum

Sei $K$ ein Körper. Ein $K$-Vektorraum ist ein Tripel $(V, +, \cdot)$ bestehend

- aus einer Menge $V$
- einer Verknüpfung $+ : V \times V \to V$ (Vektoraddition) und einer
- einer Verknüpfung $\cdot : K \times V \to V$ (skalare Multiplikation)

So dass folgendes gilt:

V1 $(V, +)$ ist eine kommutative Gruppe (auch abelsche Gruppe). Das neutrale Element 0 heißt Nullvektor und das zu $v \in V$ inverse Element $-v$ heißt der zu $v$ negative Vektor

V2 für alle $v, w \in V$ und $\lambda, \mu \in K$ gilt:
- $(\lambda + \mu) \cdot v = (\lambda \cdot v) + (\mu \cdot v)$
- $\lambda \cdot (v + w) = (\lambda \cdot v) + (\lambda \cdot w)$
- $(\lambda\mu) \cdot v = \lambda \cdot (\mu \cdot v)$
- $1 \cdot v = v$.

Die Elemente von $V$ nennt man Vektoren, die von $K$ Skalare.



## Vektorraum: Anmerkungen

- Statt $\lambda \cdot v$ mit $\lambda \in K$ und $v \in V$ schreibt man zumeist $\lambda v$
- Per Konvention bindet die skalare Multiplikation stärker als die Addition in $V$ und die Addition in $K$, also $\lambda v + w = (\lambda v) + w$
- In jedem K-Vektorraum gilt:
  - $0 \cdot v = 0$ für alle $v \in V$
  - $\lambda \cdot 0 = 0$ für alle $\lambda \in K$
  - Wenn $\lambda \in K$ und $v \in V$ und $\lambda v = 0$, dann gilt $\lambda = 0$ oder $v = 0$.
  - $(-1) \cdot v = -v$ für alle $v \in V$.

Können sie diese Aussagen beweisen?

## Beweise einiger Eigenschaften von K-Vektorräumen

Sei $(V, +, \cdot)$ ein $K$-Vektorraum.

- Aussage: $0 \cdot v = 0$ für alle $v \in V$
  Beweis: es gilt $0 \cdot v = (0+0) \cdot v = 0 \cdot v + 0 \cdot v$ und $0 \cdot v + 0 = 0 \cdot v$.
  Daraus folgt $0 \cdot v = 0$. $\square$
- Aussage: $\lambda \cdot 0 = 0$ für alle $\lambda \in K$
  Beweis: es gilt $\lambda \cdot 0 = \lambda \cdot (0+0) = \lambda \cdot 0 + \lambda \cdot 0$ und $\lambda \cdot 0 + 0 = \lambda \cdot 0$.
  Daraus folgt $\lambda \cdot 0 = 0$. $\square$
- Aussage: Wenn $\lambda \in K$ und $v \in V$ und $\lambda \cdot v = 0$, dann gilt $\lambda = 0$ oder $v = 0$.
  Beweis durch Fallunterscheidung: 1. Fall: Wenn $\lambda = 0$ und $\lambda \cdot v = 0$ dann folgt die Aussage direkt. 2. Fall: Wenn $\lambda \neq 0$ und $\lambda \cdot v = 0$, dann gilt $v = 1 \cdot v = (\lambda^{-1}\lambda) \cdot v = \lambda^{-1}(\lambda \cdot v) = \lambda^{-1} \cdot 0 = 0$. $\square$
- Aussage: $(-1) \cdot v = -v$ für alle $v \in V$.
  Beweis: aus $v + (-1) \cdot v = 1 \cdot v + (-1) \cdot v = (1-1) \cdot v = 0 \cdot v = 0$ folgt $(-1) \cdot v = -v$. $\square$



## Vektorraum Beispiel: Vektorraum der Funktionen

Setzt man $K = \mathbb{R}$ und
$V = \{f \mid f$ ist eine Funktion mit $f : \mathbb{R} \to \mathbb{R}\}$, dann bildet $(V, +, \cdot)$ einen $\mathbb{R}$-Vektorraum mit

- Vektoraddition $V \times V \to V$:
  $(f+g) : \mathbb{R} \to \mathbb{R}$ mit $(f+g)(x) = f(x) + g(x)$
- skalare Multiplikation $\mathbb{R} \times V \to V$:
  $(\lambda \cdot f) : \mathbb{R} \to \mathbb{R}$ mit $(\lambda \cdot f)(x) = \lambda f(x)$



## Vektorraum Beispiel: Koordinatenraum oder Raum der geordneten $n$-Tupel

Ist $K$ ein Körper und $n \in \mathbf{N}$, so bildet das n-fache kartesische Produkt $K^n = \{(x_1, \dots, x_n) \mid x_1, \dots, x_n \in K\}$ die Menge aller n-Tupel $(x_1, \dots, x_n)$ mit Komponenten aus $K$. $(K^n, +, \cdot)$ bildet einen $K$-Vektorraum mit

- Vektoraddition $+ : K^n \times K^n \to K^n$ definiert durch
  $$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$$
- und skalare Multiplikation $\cdot : K \times K^n \to K^n$ definiert durch
  $$a \cdot (x_1, \dots, x_n) = (a \cdot x_1, \dots, a \cdot x_n)$$

Der Vektorraum $(K^n, +, \cdot)$ wird auch als Koordinatenraum der Dimension $n$ oder als Raum der geordneten $n$-Tupel bezeichnet.



## Vektoren in Koordinatenräumen

Koordinatenvektoren notiert man häufig als Spaltenvektoren. Für das Tupel $(x_1, \dots x_n)$ schreibt man dann

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Für die Vektoraddition gilt dann:

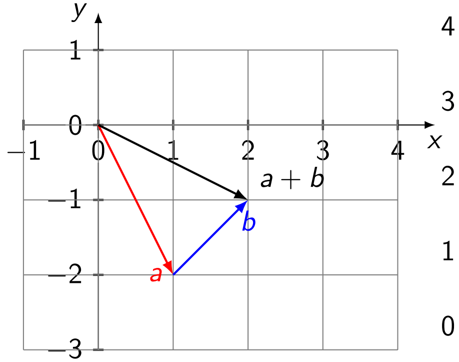
$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

und für die skalare Multiplikation gilt

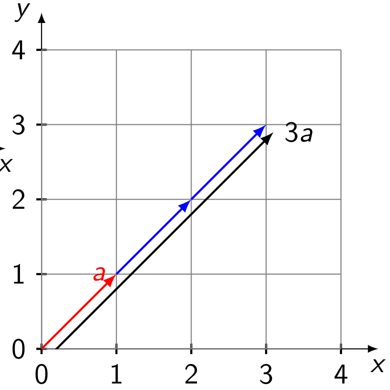
$$a \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a \cdot x_1 \\ \vdots \\ a \cdot x_n \end{pmatrix}$$

## Geometrische Interpretation von Vektoren in Koordinatenräumen

Vektoraddition:




Skalare Multiplikation:






## Skalarprodukt in $\mathbb{R}^n$

Für $v = (v_1, \ldots v_n)$ und $w = (w_1, \ldots w_n)$ mit $v, w \in \mathbb{R}^n$ ist das Skalarprodukt $\langle v, w \rangle$ definiert als

$$\langle v, w \rangle = v_1 w_1 + v_2 w_2 + \ldots v_n w_n.$$

Das Skalarprodukt zweier Vektoren ist ein Skalar:

$$\langle\ ,\ \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

Die Länge oder Norm eines Vektors $v \in \mathbb{R}^n$ ist

$$\|v\| = \sqrt{\langle v, v \rangle}$$

Die Distanz zweier Vektoren $v, w \in \mathbb{R}^n$ ist

$$d(v, w) = \|w - v\|$$



## Zusammenhang zwischen Skalarprodukt und Winkel

Seien $v, w \in \mathbb{R}^n$ und $\theta = \sphericalangle(v, w)$ der Winkel zwischen $v$ und $w$, dann gilt

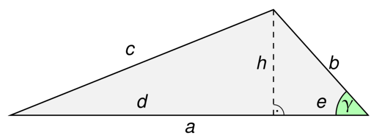
$$\langle v, w \rangle = \|v\| \|w\| \cos(\theta)$$

Für den Beweis benötigt man den Kosinussatz (`https://de.wikipedia.org/wiki/Kosinussatz`): Seien $a, b, c$ die Seiten eines Winkels und $\gamma$ der der Seite $c$ gegenüberliegende Winkel, dann gilt

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma)$$



## Beweis des Kosinussatzes




- Es gilt aufgrund des Satzes des Pythagoras:
  $h^2 = b^2 - e^2$
  $c^2 = h^2 + d^2$
- Ferner gilt:
  $d^2 = (a - e)^2 = a^2 - 2 \cdot a \cdot e + e^2$
- Daraus folgt:
  $c^2 = b^2 - e^2 + a^2 - 2 \cdot a \cdot e + e^2 = b^2 + a^2 + 2 \cdot a \cdot e$
- Nach der Definition des Kosinus gilt:
  $\cos\gamma = \frac{e}{b}$
  $e = b \cdot \cos\gamma$
- Daraus folgt:
  $c^2 = a^2 + b^2 + 2 \cdot a \cdot b \cdot \cos\gamma \square$

## Zusammenhang zwischen Skalarprodukt und Winkel

Seien $v$ und $w$ Vektoren in $\mathbb{R}^n$ und $ABC$ ein Dreieck mit den Seitenlängen $a = ||v||$, $b = ||w||$ und $c = ||v - w||$ und dem Winkel $\theta$ zwischen $v$ und $w$.
Aus dem Kosinussatz folgt:

$$\langle v - w, v - w\rangle = \langle v, v\rangle + \langle w, w\rangle - 2||v|| \cdot ||w|| \cdot \cos\theta$$

Es gilt:

$$\begin{aligned}\langle v - w, v - w\rangle &= \sum_{i=1}^{n}(v_i - w_i)^2\\ &= \sum_{i=1}^{n} v_i^2 - 2v_iw_i + w_i^2\\ &= \sum_{i=1}^{n} v_i^2 - 2\sum_{i=1}^{n} v_iw_i + \sum_{i=1}^{n} w_i^2\\ &= \langle v, v\rangle - 2\langle v, w\rangle + \langle w, w\rangle\end{aligned}$$



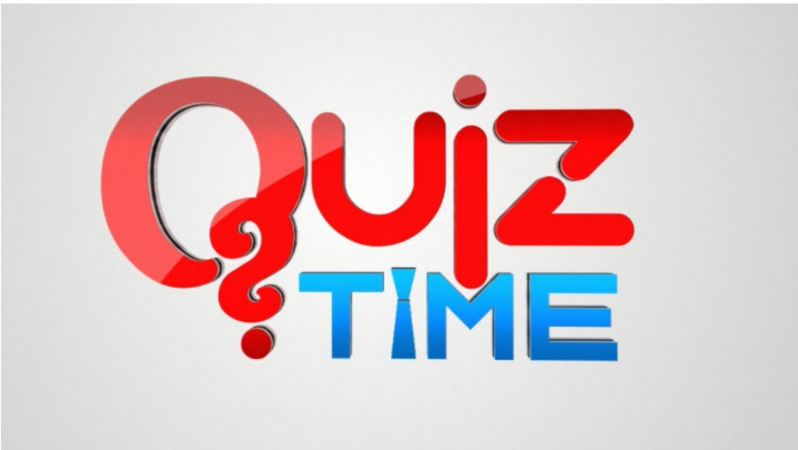
## Zusammenhang zwischen Skalarprodukt und Winkel

Daraus folgt:

$$\begin{aligned}\langle v, v\rangle - 2\langle v, w\rangle + \langle w, w\rangle &= \langle v, v\rangle + \langle w, w\rangle - 2||v|| \cdot ||w|| \cdot \cos\theta\\ -2\langle v, w\rangle &= -2||v|| \cdot ||w|| \cdot \cos\theta\\ \langle v, w\rangle &= ||v|| \cdot ||w|| \cdot \cos\theta\square\end{aligned}$$



## Quiz-Time






## Vektorraum Beispiel: Matrizenraum (1)

Ein rechteckiges Schema

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{pmatrix} = (a_{ij})_{1\leq i\leq m, 1\leq j\leq n}$$

von Elementen $a_{ij}$ aus einem Körper $K$ heißt Matrix. Die Elemente $a_{ij}$ heißen Komponenten der Matrix. Man spricht von den Zeilen und Spalten einer Matrix. Eine Matrix mit $m$ Zeilen und $n$ Spalten heißt $m \times n$-Matrix (lese: $m$ Kreuz $n$ Matrix).

## Vektorraum Beispiel: Matrizenraum (2)

Sei $K$ ein Körper und $m, n \in \mathbb{N}$. Dann ist

$$K^{m\times n} = \{(a_{ij})_{1\leq i\leq m, 1\leq j\leq n} \mid a_{ij} \in K\}$$

die Menge der Matrizen der Größe $m \times n$ mit Komponenten aus $K$. $K^{m\times n}$ bildet mit der folgenden Vektoraddition und skalaren Multiplikation einen $K$-Vektorraum:

- Vektor- oder besser Matrizenaddition:
  $+ : K^{m\times n} \times K^{m\times n} \to K^{m\times n}$ mit $A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$.
- Multiplikation mit einem Skalar:
  $\cdot : K \times K^{m\times n} \to K^{m\times n}$ mit $\lambda \cdot A = \lambda \cdot (a_{ij}) = (\lambda \cdot a_{ij})$.

Der Vektorraum $(K^{m\times n}, +, \cdot)$ wird auch der Matrizenraum oder Raum der Matrizen der Größe $m \times n$ über dem Körper $K$ genannt.

Die Vektorraumeigenschaften von $(K^{m\times n}, +, \cdot)$ folgen unmittelbar aus den Vektorraumeigenschaften von $(K^n, +, \cdot)$. Das neutrale Element der Matrizenaddition ist die Matrix $(a_{ij})$ mit $a_{ij} = 0$ für alle $i, j$. Die zu $(a_{ij}$ additiv inverse Matrix ist $(-a_{ij})$.



## Beispiel: Matrizenaddition und Multiplikation mit Skalar

Sei im folgenden $K = \mathbb{R}$.

$$\begin{pmatrix} 3 & 2 & 4 \\ 0 & 1 & -1 \end{pmatrix} + \begin{pmatrix} 1 & 3 & -2 \\ 2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3+1 & 2+3 & 4-2 \\ 0+2 & 1+0 & -1+1 \end{pmatrix} = \begin{pmatrix} 4 & 5 & 2 \\ 2 & 1 & 0 \end{pmatrix}$$

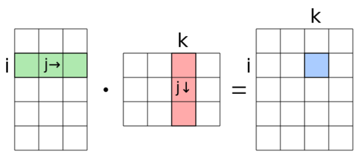
$$3 \cdot \begin{pmatrix} 3 & 2 & 4 \\ 0 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 3\cdot 3 & 3\cdot 2 & 3\cdot 4 \\ 3\cdot 0 & 3\cdot 1 & 3\cdot -1 \end{pmatrix} = \begin{pmatrix} 9 & 6 & 12 \\ 0 & 3 & -3 \end{pmatrix}$$



## Matrizenmultiplikation

Die Matrizenmultiplikation ist die Operation
$\cdot : K^{m\times p} \times K^{p\times n} \to K^{m\times n}$ mit $A \cdot B = C$ und

$$c_{ik} = \sum_{j=1}^{p} a_{ij} \cdot b_{jk}$$




Beispiel:

$$\begin{pmatrix} 3 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 4 & 0 \end{pmatrix} = \begin{pmatrix} 7 & 8 \\ 9 & 2 \end{pmatrix}$$



## transponierte Matrix

Die Transponierte einer $m \times n$-Matrix $A = (a_{ij})$ ist die $n \times m$-Matrix $A^T = (a_{ji})$. Also

$$A^T = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}^T = \begin{pmatrix} a_{11} & \dots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \dots & a_{mn} \end{pmatrix}$$

Es gelten folgende Aussagen:

- $(A + B)^T = A^T + B^T$
- $(c \cdot A)^T = c \cdot A^T$
- $(A^T)^T = A$
- $(A \cdot B)^T = B^T \cdot A^T$

Beispiel: $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$

## Beispiele: Matrizenmultiplikation

Zusammenhang zwischen Skalarprodukt von Vektoren und Matrizenmultiplikation:

$$\left\langle \begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}-1\\0\\1\end{pmatrix} \right\rangle = 2 \in K$$

$$\begin{pmatrix}1\\2\\3\end{pmatrix}^T \cdot \begin{pmatrix}-1\\0\\1\end{pmatrix} = \begin{pmatrix}1 & 2 & 3\end{pmatrix} \cdot \begin{pmatrix}-1\\0\\1\end{pmatrix} = \begin{pmatrix}2\end{pmatrix} \in K^{1\times 1}$$

Multiplikation einer Matrix und eines Vektors

$$\begin{pmatrix}1 & 4\\2 & 5\\3 & 6\end{pmatrix} \cdot \begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}6\\9\\12\end{pmatrix}$$



## Matrizen als lineare Abbildungen

Seien $V$ und $W$ Vektorräume. Eine Abbildung $f : V \to W$ heißt linear oder ein Homomorphismus, wenn folgendes gilt:

1. $f(x + y) = f(x) + f(y)$ und
2. $f(\lambda x) = \lambda f(x)$.

Bemerkungen:

- Jede Matrix $A \in K^{m\times n}$ definiert eine lineare Abbildung $A : K^n \to K^m$ mit $x \mapsto Ax$.
- Die Matrizenmultiplikation entspricht der Komposition der linearen Abbildungen: Sei $A \in K^{m\times n}$ und $B \in K^{r\times m}$, dann ist $B \circ A : K^n \to K^r$ mit $B \circ A(x) = (B \cdot A)x$.

Beispiel:

$$\left(\begin{pmatrix}1 & 2 & 3\\4 & 5 & 6\end{pmatrix} \cdot \begin{pmatrix}1 & 4 & 1 & 0\\2 & 5 & 0 & 1\\3 & 6 & 2 & 1\end{pmatrix}\right) \begin{pmatrix}1\\2\\2\\1\end{pmatrix} = \begin{pmatrix}14 & 32 & 7 & 5\\32 & 77 & 16 & 11\end{pmatrix} \begin{pmatrix}1\\2\\2\\1\end{pmatrix} = \begin{pmatrix}97\\229\end{pmatrix}$$

$$\begin{pmatrix}1 & 2 & 3\\4 & 5 & 6\end{pmatrix} \circ \begin{pmatrix}1 & 4 & 1 & 0\\2 & 5 & 0 & 1\\3 & 6 & 2 & 1\end{pmatrix} \left(\begin{pmatrix}1\\2\\2\\1\end{pmatrix}\right) = \begin{pmatrix}1 & 2 & 3\\4 & 5 & 6\end{pmatrix} \left(\begin{pmatrix}11\\13\\20\end{pmatrix}\right) = \begin{pmatrix}97\\229\end{pmatrix}$$



## Matrizen zur Beschreibung linearer Gleichungssysteme

Ein Gleichungssystem mit n Variablen kann man auch als Matrixproblem beschreiben:
Beispiel: Das Gleichungssystems

$$\begin{array}{rcrcrcr} -x & + & 2y & + & z & = & -2\\ 3x & - & 8y & - & 2z & = & 4\\ x & + & & & 4z & = & -2 \end{array}$$

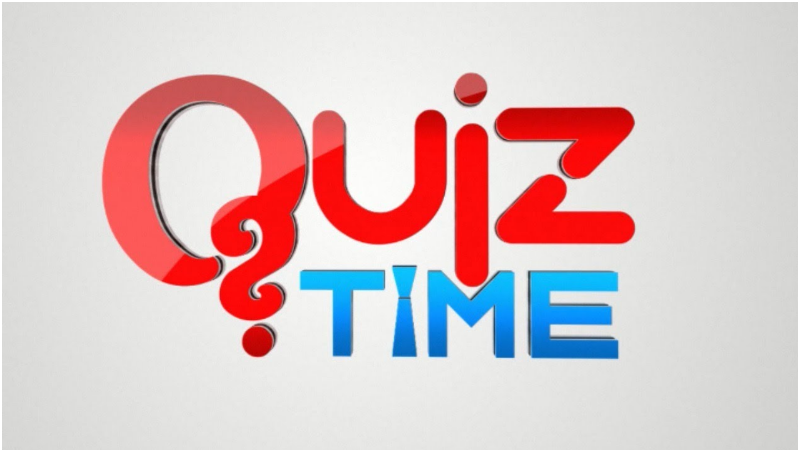
lässt sich beschreiben als

$$\begin{pmatrix}-1 & 2 & 1\\3 & -8 & -2\\1 & 0 & 4\end{pmatrix} \begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}-2\\4\\-2\end{pmatrix}$$

Zur Lösung solcher Systeme kann z.B. des Gaußsche Eliminationsverfahren eingesetzt werden.
(Lösung für das Beispiel $x = 2, y = 0.5, z = -1$).



## Quiz-Time



## Distributionelle Semantik

Die folgenden Folien entstammen dem Foliensatz "Distributional Semantics" von Marco Baroni und Gemma Boleda zu dem Kurs "CS 388: Natural Language Processing".
`https://www.cs.utexas.edu/~mooney/cs388/slides/dist-sem-intro-NLP-class-UT.pdf`



## The distributional hypothesis

- The meaning of a word is the set of contexts in which it occurs in texts
- *Important aspects of* the meaning of a word *are a function of* (*can be approximated by*) the set of contexts in which it occurs in texts



## The distributional hypothesis in real life

McDonald & Ramscar 2001

He filled the wampimuk, passed it around and we all drunk some

We found a little, hairy wampimuk sleeping behind the tree



## The distributional hypothesis, weak and strong

Lenci (2008)

- Weak: a quantitative method for semantic analysis and lexical resource induction
- Strong: A cognitive hypothesis about the form and origin of semantic representations

## Advantages of distributional semantics

Distributional semantic models are

- model of inductive *learning* for word meaning
- radically empirical
- rich
- flexible
- cheap, scalable



## Constructing the models

- Pre-process the source corpus
- Collect a co-occurrence matrix (with *distributional vectors* representing words as rows, and contextual elements of some kind as columns/dimensions)
- Transform the matrix: re-weighting raw frequencies, dimensionality reduction
- Use resulting matrix to compute word-to-word similarity



## Distributional vectors

- Count how many times each target word occurs in a certain context
- Build vectors out of (a function of) these context occurrence counts
- Similar words will have similar vectors



## Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

| | |
|---|---|
| bark | ++ |
| park | + |
| owner | + |
| leash | + |

## Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

| | |
|---|---|
| bark | + |
| park | |
| owner | |
| leash | |



## Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

| | |
|---|---|
| bark | + |
| park | + |
| owner | |
| leash | |



## Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

| | |
|---|---|
| bark | + |
| park | + |
| owner | + |
| leash | |



## Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

| | |
|---|---|
| bark | + |
| park | + |
| owner | + |
| leash | + |

## Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

| | |
|---|---|
| bark | ++ |
| park | + |
| owner | + |
| leash | + |



## The co-occurrence matrix

| | leash | walk | run | owner | pet | bark |
|---|---|---|---|---|---|---|
| dog | 3 | 5 | 2 | 5 | 3 | 2 |
| cat | 0 | 3 | 3 | 2 | 3 | 0 |
| lion | 0 | 3 | 2 | 0 | 1 | 0 |
| light | 0 | 0 | 0 | 0 | 0 | 0 |
| bark | 1 | 0 | 0 | 2 | 1 | 0 |
| car | 0 | 0 | 1 | 3 | 0 | 0 |



## What is “context”?

DOC1: The silhouette of the sun beyond a wide-open bay on the lake; the sun still glitters although evening has arrived in Kuhmo. It’s midsummer; the living room has its instruments and other objects in each of its corners.



## What is “context”?

Documents

DOC1: The silhouette of the sun beyond a wide-open bay on the lake; the sun still glitters although evening has arrived in Kuhmo. It’s midsummer; the living room has its instruments and other objects in each of its corners.

## What is "context"?

All words in a wide window

DOC1: The silhouette of the sun beyond a wide-open bay on the lake; the sun still glitters although evening has arrived in Kuhmo. It's midsummer; the living room has its instruments and other objects in each of its corners.



## What is "context"?

Content words only

DOC1: The silhouette of the sun beyond a wide-open bay on the lake; the sun still glitters although evening has arrived in Kuhmo. It's midsummer; the living room has its instruments and other objects in each of its corners.



## What is "context"?

Content words in a narrower window

DOC1: The silhouette of the sun beyond a wide-open bay on the lake; the sun still glitters although evening has arrived in Kuhmo. It's midsummer; the living room has its instruments and other objects in each of its corners.



## What is "context"?

POS-coded content lemmas

DOC1: The silhouette-n of the sun beyond a wide-open-a bay-n on the lake-n; the sun still glitter-v although evening-n has arrive-v in Kuhmo. It's midsummer; the living room has its instruments and other objects in each of its corners.

## What is “context”?

POS-coded content lemmas filtered by syntactic path to the target

DOC1: The silhouette-n of the sun beyond a wide-open bay on the lake; the sun still glitter-v although evening has arrived in Kuhmo. It’s midsummer; the living room has its instruments and other objects in each of its corners.



## What is “context”?

. . . with the syntactic path encoded as part of the context

DOC1: The silhouette-n_ppdep of the sun beyond a wide-open bay on the lake; the sun still glitter-v_subj although evening has arrived in Kuhmo. It’s midsummer; the living room has its instruments and other objects in each of its corners.

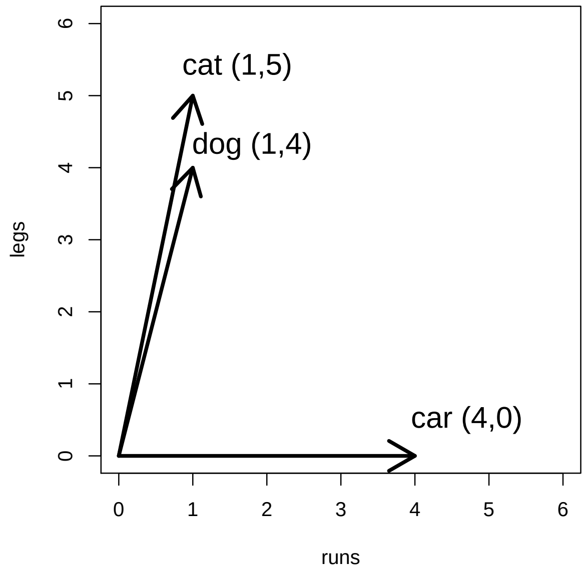


## Contexts as vectors

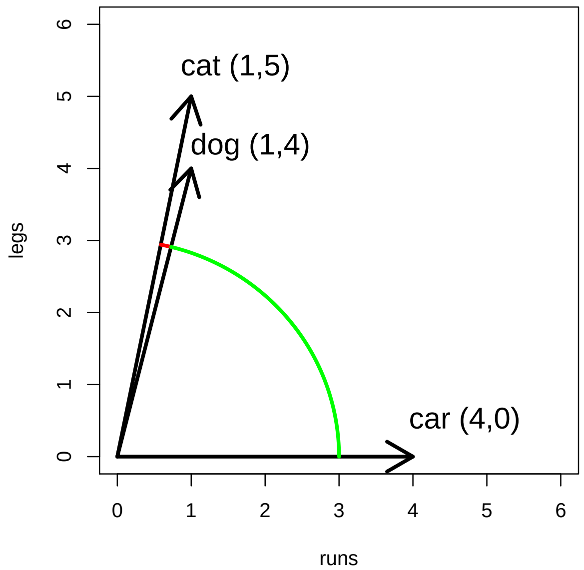
| | runs | legs |
|---|---|---|
| dog | 1 | 4 |
| cat | 1 | 5 |
| car | 4 | 0 |



## Semantic space



## Semantic similarity as angle between vectors






## Computing the cosine

### Example

$$\frac{\sum_{i=1}^{i=n} a_i \times b_i}{\sqrt{\sum_{i=1}^{i=n} a_i^2} \times \sqrt{\sum_{i=1}^{i=n} b_i^2}}$$

| | runs | legs |
|---|---|---|
| dog | 1 | 4 |
| cat | 1 | 5 |
| car | 4 | 0 |

$$\text{cosine}(\textit{dog},\textit{cat}) = \frac{(1\times 1)+(4\times 5)}{\sqrt{1^2+4^2}\times\sqrt{1^2+5^2}} = 0.9988681$$

arc-cosine(0.9988681) = 2.72 degrees

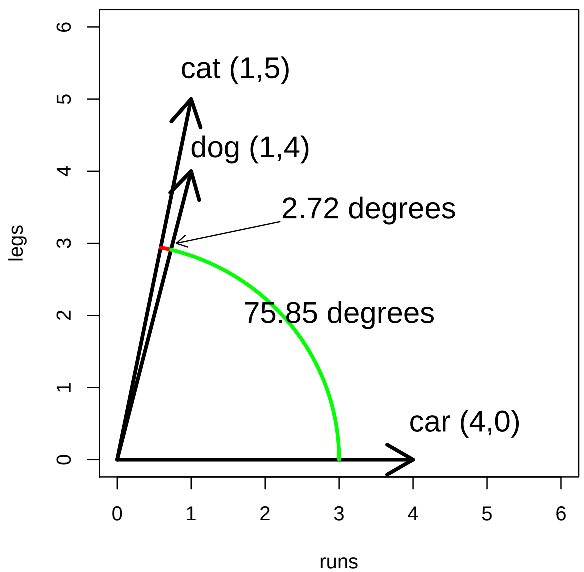
$$\text{cosine}(\textit{dog},\textit{car}) = \frac{(1\times 4)+(4\times 0)}{\sqrt{1^2+4^2}\times\sqrt{4^2+0^2}} = 0.2425356$$

arc-cosine(0.2425356) = 75.85 degrees



## Computing the cosine

### Example






## Nearest neighbour examples

### BNC, 2-content-word-window context

| rhino | fall | rock |
|---|---|---|
| woodpecker | rise | lava |
| rhinoceros | increase | sand |
| swan | fluctuation | boulder |
| whale | drop | ice |
| ivory | decrease | jazz |
| plover | reduction | slab |
| elephant | logarithm | cliff |
| bear | decline | pop |
| satin | cut | basalt |
| sweatshirt | hike | crevice |

## Nearest neighbour examples

BNC, 2-content-word-window context

| green | good | sing |
|---|---|---|
| blue | bad | dance |
| yellow | excellent | whistle |
| brown | superb | mime |
| bright | poor | shout |
| emerald | improved | sound |
| grey | perfect | listen |
| speckled | clever | recite |
| greenish | terrific | play |
| purple | lucky | hear |
| gleaming | smashing | hiss |



# Mathematische Grundlagen der Computerlinguistik

## Graphentheorie

Dozentin: Wiebke Petersen

10. Foliensatz



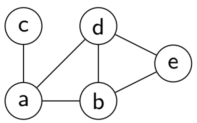
## Graph

Ein Graph ist ein geordnetes Paar $(V, E)$ bestehend aus einer Menge $V$ und einer Menge $E$ von zweielementigen Teilmengen von $V$. Es gilt $V \cap E = \emptyset$.

Die Elemente von $V$ heißen Ecken (vertices) und die von $E$ Kanten (edges).

- Statt von den Ecken spricht man auch oft von den Knoten eines Graphen
- Die hier definierten Graphen werden häufig auch einfache Graphen genannt.
- Graphen mit endlicher Eckenmenge heißen endliche Graphen. Sie lassen sich durch Diagramme visualisieren:

$V = \{a, b, c, d, e\}, E = \{\{a, b\}, \{a, c\}, \{a, d\}, \{b, e\}, \{b, d\}, \{d, e\}\}$

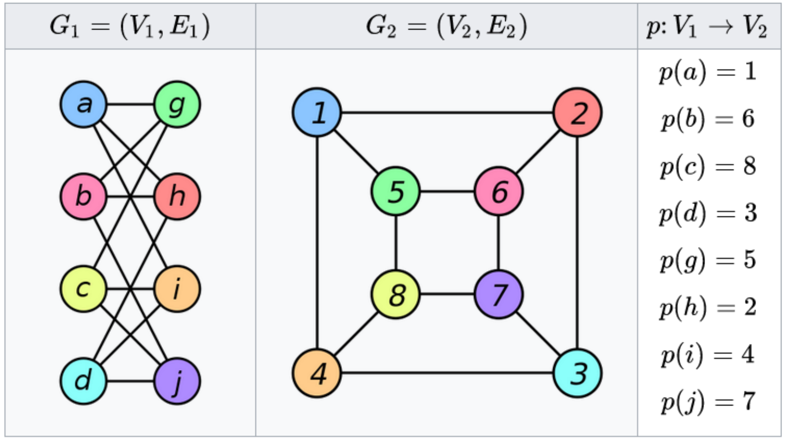





## Terminologie (Graphen)

- Die Ordnung eines Graphen $(V, E)$ ist $|V|$. Graphen mit Ordnungen kleiner 2 heißen trivial.
- Eine Ecke $v$ heißt mit einer Kante $e$ inzident, wenn $v \in e$.
- Der Grad einer Ecke $v$ ist die Zahl der mit $v$ inzidenten Kanten.
- Zwei Ecken $v_1, v_2 \in V$ heißen adjazent oder benachbart, wenn $\{v_1, v_2\} \in E$.
- Ein Graph, in dem je zwei Ecken benachbart sind, heißt vollständig.
- Zwei Graphen $(V, E)$ und $(V', E')$ heißen isomorph, wenn es eine Bijektion $\phi : V \to V'$ gibt mit: $\{v, w\} \in E \Leftrightarrow \{\phi(v), \phi(w)\} \in E'$.
- $(V', E')$ ist ein Teilgraph von $(V, E)$, wenn $V' \subseteq V$ und $E' \subseteq E$.

## Beispiel für isomorphe Graphen




https://de.wikipedia.org/wiki/Isomorphie_von_Graphen



## Graphen und symmetrische Relationen

- Jeder Graph $(V, E)$, definiert eine binäre, symmetrische, irreflexive Relation $E$ auf $V$.
- Jede binäre, symmetrische, irreflexive Relation $R'$ auf einer Menge $M$ definiert einen Graphen $(M, R)$.

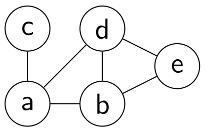
Frage: Warum muss die Relation symmetrisch und irreflexiv sein, um einen Graphen zu definieren?



## Wege in Graphen

Sei $(V, E)$ ein Graph.

- Eine Folge von Knoten $(v_1, v_2, \dots, v_n)$ bildet einen Weg in $(V, E)$, genau dann wenn für alle $1 \leq i < n$ gilt: $\{v_i, v_{i+1}\} \in E$ und wenn $v_i \neq v_j$ für alle $i \neq j$.
- Ein Folge von Knoten $(v_1, v_2, \dots, v_n)$ in $(V, E)$ heißt Kreis, wenn $v_1 = v_n$, $\{v_{n-1}, v_n\} \in E$ und $(v_1, v_2, \dots, v_{n-1})$ einen Weg in $(V, E)$ bildet.






## Terminologie (Wege)

- Wenn $(v_1, v_2, \dots, v_n)$ ein Weg in $(V, E)$ ist, dann heißen $v_1$ und $v_n$ die Endknoten des Weges.
- Knoten eines Wegs, die keine Endknoten sind, heißen innere Knoten.
- Wege in Graphen können auch als Teilgraphen des Graphen aufgefasst werden.
- Ein Knoten $u$ heißt von einem Knoten $v$ erreichbar in einem Graphen, wenn es einen Weg gibt, der $v$ und $u$ als Endknoten hat. Sonst heißt $u$ unerreichbar von $v$.
- Ein Graph in dem für je zwei beliebige Knoten gilt, dass sie untereinander erreichbar sind, heißt zusammenhängend.
- Zwei Wege heißen kreuzungsfrei oder (knoten)disjunkt, wenn sie keine inneren Knoten gemeinsam haben.
- Die Länge eines Weges $(v_1, v_2, \dots, v_n)$ ist die Zahl seiner Kanten, sprich $n - 1$.

## Terminologie (Kreise)

- Ein Kreis $(v_1, v_2, v_3, v_1)$ aus drei Kanten heißt Dreieck
- Ein Graph heißt zyklisch, wenn er mindestens einen Kreis enthält, sonst heißt er azyklisch



## Bäume und Wälder

Ein kreisfreier Graph heißt Wald.

Ein zusammenhängender Wald heißt Baum.

Die Ecken vom Grad 1 eines Baums heißen Blätter.

Satz:
Jeder zusammenhängende Graph wird von einem Baum aufgespannt. Ein Baum $(V', E')$ spannt einen Graphen $(V, E)$ auf, wenn $V' = V$ und $E' \subseteq E$. Dieser Baum wird auch Spannbaum des Graphen genannt.



## Planare Graphen

Ein Graph heißt planar oder plättbar, wenn er auf einer Ebene mit Punkten für die Knoten und Linien für die Kanten dargestellt werden kann, sodass sich keine Kanten schneiden.

**Hinweis:** Jeder Baum ist plättbar. Warum?

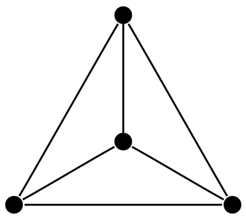


Abbildung: Planare Zeichnung des $K_4$

Quelle: `https://de.wikipedia.org/wiki/Planarer_Graph`



## Berühmte Probleme der Graphentheorie

Die Graphentheorie ist ein extrem breites Gebiet mit zahlreichen Anwendungen, wichtigen Algorithmen und bekannten Problemen. Hier ein kleiner Appetizer.

- Das Königsberger Brückenproblem (`http://www.matheprisma.uni-wuppertal.de/Module/Koenigsb/index.htm`)
- Das Vierfarbenproblem (`http://www.matheprisma.uni-wuppertal.de/Module/4FP/index.htm`)
- Suche kürzester Wege (`http://www.matheprisma.uni-wuppertal.de/Module/Graphen/index.htm`)

Im folgenden werden einige häufig verwendete Erweiterungen einfacher Graphen eingeführt.

## Gerichtete Graphen

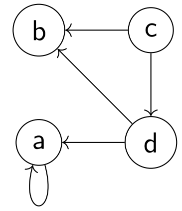
Ein gerichteter Graph (auch Digraph, 'directed graph') besteht aus einer Menge V von Knoten und einer Menge geordneter Knotenpaare $E \subseteq V \times V$ von Kanten.

- Die Kanten $(v, w) \in E$ eines gerichteten Graphen sind gerichtete Kanten. Die Darstellung erfolgt meistens als Pfeil. Dieser gibt die zu durchlaufende Richtung an.
- Digraphen werden zum Beispiel zur Darstellung von endlichen Automaten verwendet.
- Eine gerichtete Kante $e = (x, y)$ geht von x nach y. Wobei x der Startknoten und y der Endknoten von e ist. Außerdem gilt y als der direkte Nachfolger von x und x als direkter Vorgänger von y.



## Beispiel gerichteter Graph

$$V = \{a, b, c, d\},\ E = \{(d, b), (d, a), (c, d), (c, b), (a, a)\}$$






## Multigraphen

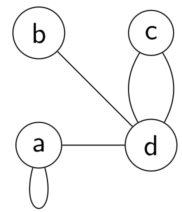
Ein Multigraph besteht aus einer Menge V von Knoten, einer Menge E von Kanten und einer Funktion $f : E \to \{\{x, y\} : x, y \in V\}$, welche jeder Kante einen oder zwei Endecken zuweist.

- In einem Multigraph können zwei Knoten durch mehrere, unterschiedene Kanten miteinander verbunden sein. Hierbei spricht man auch von Mehrfachkanten.
- Kanten, die von einer Ecke zu dieser zurück laufen (also eine Ecke mit sich selbst verbinden), werden Schlingen genannt.



## Beispiel Multigraph

$$V = \{a, b, c, d\},\ E = \{e_1, e_2, e_3, e_4, e_5\}$$
$$e_1 \mapsto \{d, c\},\ e_2 \mapsto \{d, c\},\ e_3 \mapsto \{d, a\},\ e_4 \mapsto \{d, b\},\ e_5 \mapsto \{a\}$$

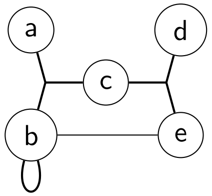


## Hypergraph

Ein Hypergraph ist ein Tupel (V,E), wobei $V = \{v_1, v_2, \dots, v_n\}$ die Eckenmenge und $E \subseteq \mathcal{P}(V)$ mit $\emptyset \notin E$ die Menge der Hyperkanten bezeichnet.

- Eine Kante (bzw. Hyperkante) verbindet hier also nicht nur zwei, sondern mehrere Knoten gleichzeitig.

$$V = \{a, b, c, d, e\},\ E = \{\{a, b, c\}, \{c, d, e\}, \{b, e\}, \{b\}\}$$






## Gewichtete Graphen

Ein Kantengewicht wird durch die Kantengewichtsfunktion f: E $\rightarrow \mathbb{R}$ gegeben. Diese Funktion ordnet jeder Kante eine reelle Zahl als Gewicht zu. Hierbei wird das Kantengewicht einer Kante $e \in E$ mit f(e) oder $f_e$ bezeichnet.

Genauso kann jedem Knoten ein Knotengewicht gegeben werden (sprich: Jedem Knoten wird eine reele Zahl als Gewicht zugeordnet).

- Zu einem knoten-/kantengewichteten Graphen gehört also neben der Angabe der Knoten- und Kantenmenge auch die Angabe einer Funktion, die von den Knoten/Kanten in die Menge der reellen Zahlen abbildet.
- Mit Kantengewichten kann man z.B. die Stärke der Bindung zwischen zwei Knoten modellieren.
- Knotengewichte können die Wichtigkeit eines Knotens modellieren.



## Repräsentation eines Graphen als Adjazenzmatrix

Eine Adjazenzmatrix eines Graphen ist eine Matrix, die speichert, welche Knoten des Graphen durch eine Kante verbunden sind. Sie besitzt für jeden Knoten eine Zeile und eine Spalte, woraus sich für n Knoten eine n×n-Matrix ergibt. Ein Eintrag in der i-ten Zeile und j-ten Spalte gibt hierbei an, ob der i-te und der j-te Knoten adjazent sind. Steht an dieser Stelle eine 0, ist keine Kante vorhanden – eine 1 gibt an, dass eine Kante existiert.

Mit Adjazenzmatrizen lassen sich einfache Graphen, gerichtete Graphen und kantengewichtete Graphen repräsentieren.

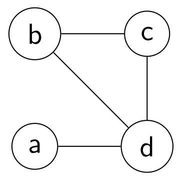


## Adjazenzmatrix: einfache Graphen

Sei $G = (V, E)$ ein einfacher Graph. Die Adjazenzmatrix $A_G = [a_{ij}]$ des Graphen G ist durch seine Einträge definiert als:

$$a_{ij} = \begin{cases} 1, & \text{falls } \{i, j\} \in E, \\ 0, & \text{sonst} \end{cases}$$

Beispiel:




| | a | b | c | d |
|---|---|---|---|---|
| a | 0 | 0 | 0 | 1 |
| b | 0 | 0 | 1 | 1 |
| c | 0 | 1 | 0 | 1 |
| d | 1 | 1 | 1 | 0 |

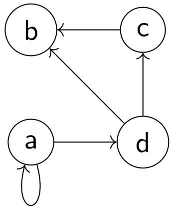
$$A_G = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

## Adjazenzmatrix: gerichtete Graphen

Sei $G = (V, E)$ ein gerichteter Graph. Die Adjazenzmatrix $A_G = [a_{ij}]$ des Graphen G ist durch seine Einträge definiert als:

$$a_{ij} = \begin{cases} 1, & \text{falls } (i,j) \in E, \\ 0, & \text{sonst} \end{cases}$$

Beispiel:




$$A_G = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

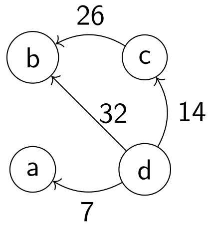


## Adjazenzmatrix: kantengewichtete Graphen

Sei $G = (V, E, f)$ ein gerichteter Graph. Die Adjazenzmatrix $A_G = [a_{ij}]$ des kantengewichteten Graphen $G = (V, E, f)$ ist durch seine Einträge definiert als

$$a_{ij} = \begin{cases} f_{(i,j)}, & \text{falls } (i,j) \in E, \\ 0, & \text{sonst} \end{cases}$$

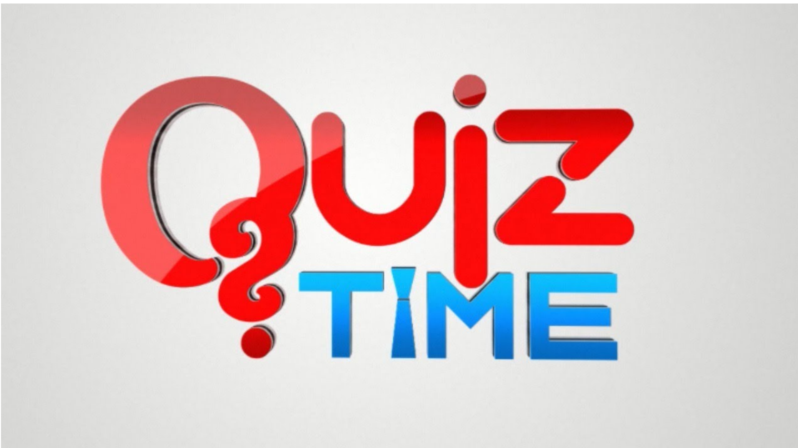
Beispiel:




$$A_G = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 26 & 0 & 0 \\ 7 & 32 & 14 & 0 \end{pmatrix}$$



## Quiz-Time





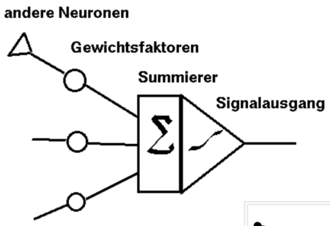

## Anwendungsgebiete in der Computerlinguistik

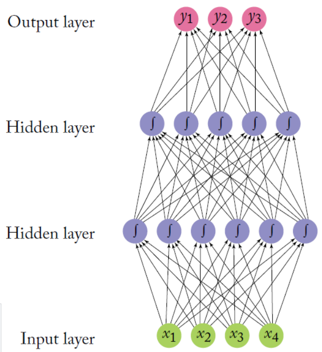
Einige spezielle Formen von Graphen haben wir bereits in den letzten Wochen kennengelernt:

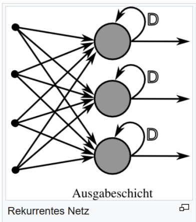
- endliche Automaten für formale Sprachen
- Hassediagramme für geordnete Mengen
- Ableitungsbäume

Weitere Beispiele auf den kommenden Folien

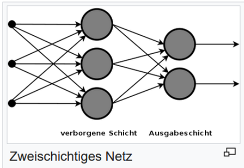
# Neuronale Netze



Rekurrentes Netz



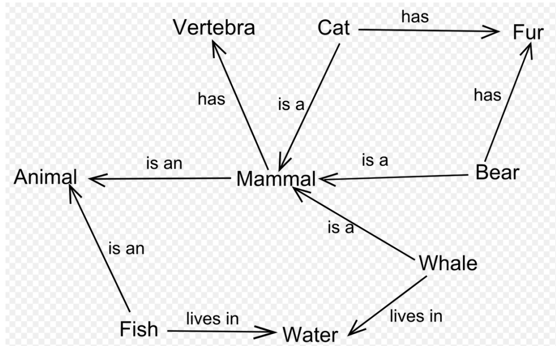
Zweischichtiges Netz

gerichtete Graphen mit gewichteten Kanten (und Schlingen)

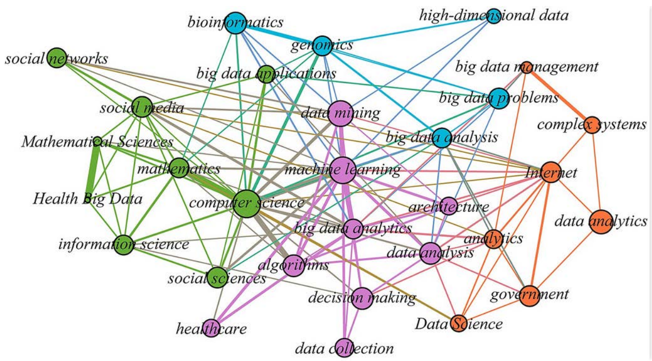
# Semantische Netze

gerichtete Graphen mit gelabelten Kanten

Vereinigungsgraph mehrerer Relationen (hier: is_a, has, lives_in)



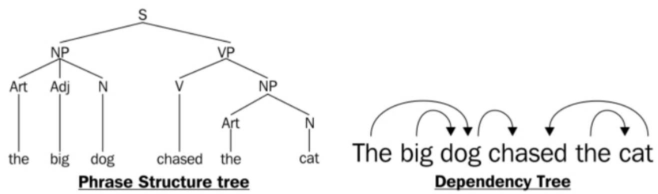

# Co-occurrence Netze




ungerichtete Graphen mit gewichteten Kanten

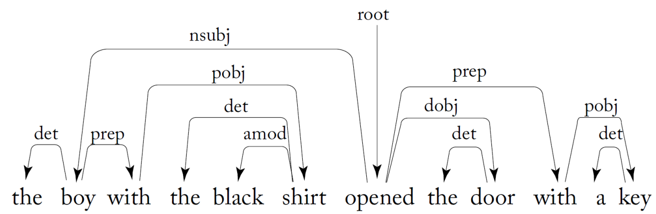
Im Beispiel: Gewichtung über Kantendicke

# Syntaktische Bäume




gerichtete Bäume mit gelabelten Knoten und geordneten Blättern

Bei Dependenzbäumen auch gelabelte Kanten



## Mathematische Grundlagen der Computerlinguistik

## Beweisverfahren

Dozentin: Wiebke Petersen

11. Foliensatz



## direkter / konstruktiver Beweis

Die Aussage wird durch die Überführung der Prämisse in die Konklusion (oder der linken Gleichungsseite in die rechte) mithilfe erlaubter Transformationen bzw. logischer Schlüsse bewiesen.

### Satz

*Wenn eine Zahl größer als 7 ist, dann ist sie auch größer als 5.*

- Sei $n$ eine beliebige Zahl größer 7 ($n > 7$).
- Es gilt außerdem $7 > 5$.
- Wegen der Transitivität der Ordnungsrelation $>$ folgt aus $n > 7$ und $7 > 5$, dass $n > 5$.
- Somit folgt, dass jede Zahl, die größer als 7 ist auch größer als 5 ist.



## direkter / konstruktiver Beweis

### Satz

*Für eine endliche Menge M gilt: Wenn $|M| = n$, dann $|\mathcal{POT}(M)| = 2^n$*

- Man zeigt: $|\mathcal{POT}(M)| = \left|\{w \in \{0,1\}^* : |w| = n\}\right|$.
- $\left|\{w \in \{0,1\}^* : |w| = n\}\right|$ ist gleich der Anzahl der Möglichkeiten aus einer Menge mit 2 Elementen $n$-mal mit Zurücklegen und unter Beachtung der Reihenfolge ein Element zu ziehen.
- Hierfür ist die Anzahl bekannt, nämlich $2^n$.



## direkter / konstruktiver Beweis

### Satz

*Die Summe der ersten n natürlichen Zahlen ist $\frac{n\cdot(n+1)}{2}$. Es gilt also:*

$$\sum_{i=1}^{i=n} i = \frac{n\cdot(n+1)}{2}$$

- Man schreibe die $n$-Zahlen zweimal nebeneinander auf und zwar einmal in aufsteigender und darunter in absteigender Reihenfolge:

| | | | | |
|---|---|---|---|---|
| 1 | 2 | ... | n-1 | n |
| n | n-1 | ... | 2 | 1 |

- Wenn man nun die Spalten zusammenrechnet, so erhält man für jede Spalte $n+1$:

$$\begin{array}{ccccc} 1 & 2 & \ldots & n-1 & n \\ n & n-1 & \ldots & 2 & 1 \\ \hline n+1 & n+1 & \ldots & n+1 & n+1 \end{array}$$

- Die Werte in der letzten Zeile zusammengezählt ergeben $n\cdot(n+1)$. Da in der letzten Zeile spaltenweise die beiden oberen Zeilen addiert worden sind, gilt:

$$2\cdot\sum_{i=1}^{i=n} i = n\cdot(n+1) \text{ folglich gilt: } \sum_{i=1}^{i=n} i = \frac{n\cdot(n+1)}{2}$$

## Beweis durch Kontraposition

- Logisches Prinzip: $(p \rightarrow q) \leftrightarrow (\neg q \rightarrow \neg p)$
- Um eine Schlussfolgerung zu beweisen, wird gezeigt, dass aus der Falschheit der Folgerung die Falschheit der Voraussetzung folgt

**Satz**

*Sei $n \in \mathbb{Z}$. Wenn $n^2$ gerade ist, ist auch $n$ gerade.*

- Angenommen, $n$ ist ungerade, also gibt es ein $k \in \mathbb{Z}$, sodass $n = 2k + 1$
- Dann ist $n^2 = (2k+1)^2 = 4k^2 + 4k + 1$
- Da $4k^2$ und $4k$ gerade Zahlen sind, ist $n^2 = 4k^2 + 4k + 1$ ungerade.



## indirekter Beweis / Beweis durch Widerspruch

Spezialfall der Kontraposition: $(\top \rightarrow p) \leftrightarrow (\neg p \rightarrow \bot)$
Wir nehmen an, dass die Aussage nicht stimmt und zeigen, dass diese Annahme zu einem logischen Widerspruch führt, also zu einer Aussage, die zugleich wahr und falsch sein muss.



## indirekter Beweis / Beweis durch Widerspruch

**Satz**

*Es gibt unendlich viele Primzahlen.*

- Wir nehmen an, es gäbe nur endlich viele Primzahlen $p_1, p_2, \dots p_n$ und $p_n$ wäre die größte aller Primzahlen.
- Dann ist
$$p = 1 + p_1 \cdot p_2 \cdot \ldots \cdot p_n$$
ebenfalls eine Primzahl, da $p$ bei der Division durch jede der Primzahlen $p_1, p_2, \dots p_n$ den Rest 1 ergibt und somit $p$ durch keine dieser Primzahlen teilbar ist.
- Zusätzlich muss aber auch $p > p_n$ gelten, woraus folgt, dass $p$ keine Primzahl ist, da $p$ größer als die größte Primzahl ist.
- Dies führt zu einem Widerspruch, da die Aussagen "$p$ ist eine Primzahl" und "$p$ ist keine Primzahl" nicht beide zugleich wahr sein können.



## indirekter Beweis / Beweis durch Widerspruch

**Satz**

*Die Potenzmenge $\mathcal{POT}(M)$ einer (abzählbaren) Menge $M$ ist immer mächtiger als die Menge selbst.*

1. $\mathcal{POT}(M)$ ist mindestens so mächtig wie $M$, da die Menge der Einermengen $\{\{m\} : m \in M\}$ genauso mächtig ist wie $M$ und eine echte Teilmenge von $\mathcal{POT}(M)$ ist.
2. Über das (generalisierte) Diagonalverfahren zeigt man, dass die Annahme, $\mathcal{POT}(M)$ und $M$ seien gleichmächtig, zu einem Widerspruch führt.
3. Aus 1 und 2 folgt, daß $\mathcal{POT}(M)$ mächtiger als $M$ sein muss.

## Beweis durch Gegenbeispiel

Um die Falschheit einer Aussage zu zeigen, genügt es ein Gegenbeispiel anzugeben.

**Satz**

*Die Schnittmenge zweier Mengen $A$ und $B$ ist nicht notwendig leer.*

Wir zeigen dass die Aussage: "Die Schnittmenge zweier Mengen $A$ und $B$ ist immer leer" falsch ist.
Gegenbeispiel: Wenn $A = \{1, 2\}$ und $B = \{1\}$, dann $A \cap B = \{1\} \neq \emptyset$.



## Ringschluss

Seien $p_1, \ldots, p_n$ Aussagen, deren Äquivalenz zu beweisen ist.
Dann genügt es zu zeigen, dass für $k < n$ gilt $p_k \to p_{k+1}$ und $p_n \to p_1$
Das Verfahren basiert auf der Transitivität der logischen Schlussfolgerung $(((p \to q) \wedge (q \to r)) \to (p \to r))$

$$\begin{tikzcd}
& p_1 \arrow[dr] & \\
p_3 \arrow[ur] & & p_2 \arrow[ll]
\end{tikzcd}$$



## Ringschluss

**Satz**

*Für einen Vektor $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ sind äquivalent:*

1. $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
2. *Für jeden anderen Vektor* $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2$ *ist* $x_1 \cdot y_1 + x_2 \cdot y_2 = 0$
3. $\sqrt{x_1^2 + x_2^2} = 0$

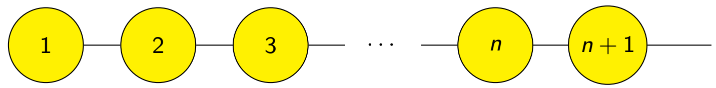
- $1 \to 2$: Nach Annahme ist $x_1 = x_2 = 0$, also $x_1 \cdot y_1 + x_2 \cdot y_2 = 0 + 0 = 0$
- $2 \to 3$: Mit $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} := \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ folgt aus 2, dass
  $0 = \sqrt{0} = \sqrt{x_1 \cdot y_1 + x_2 \cdot y_2} = \sqrt{x_1^2 + x_2^2}$
- $3 \to 1$: Nach Annahme ist $x_1^2 + x_2^2 = 0$. Da $x_1^2 \geq 0$ und $x_2^2 \geq 0$, muss $x_1^2 = x_2^2 = 0$, also $x_1 = x_2 = 0$.



## Beweis durch vollständige Induktion

In einem vollständigen Induktionsbeweis macht man sich eine besondere Eigenschaft der Menge der natürlichen Zahlen zunutze: Ausgehend von der Zahl 1 kann jede natürliche Zahl durch wiederholtes Anwenden der Nachfolgefunktion $(f(n) = n + 1)$ erreicht werden.

- Man zeigt zunächst, dass die zu beweisende Aussage für $n = 1$ gilt.
- Dann zeigt man, dass immer wenn die Aussage für ein beliebiges $n$ gilt, gilt sie auch für $n + 1$.
- Wenn die Aussage für $n = 1$ gilt und wenn außerdem aus der Gültigkeit der Aussage für $n$ auch die Gültigkeit der Aussage für $n + 1$ folgt, so folgt, dass die Aussage für alle $n \in \mathbb{N}$ gilt.



# Beweis durch vollständige Induktion

## Satz

*Die Summe der ersten $n$ natürlichen Zahlen ist $\frac{n\cdot(n+1)}{2}$. Es gilt also:*

$$\sum_{i=1}^{n} i = \frac{n\cdot(n+1)}{2}$$

**Induktionsanfang.** Für $n = 1$ ist die Summe der ersten $n$ natürlichen Zahlen 1. Da $\frac{1\cdot(1+1)}{2} = 1$ gilt die Aussage für $n = 1$.



# Beweis durch vollständige Induktion

## Satz

*Die Summe der ersten $n$ natürlichen Zahlen ist $\frac{n\cdot(n+1)}{2}$. Es gilt also:*

$$\sum_{i=1}^{n} i = \frac{n\cdot(n+1)}{2}$$

**Induktionsschluss.** Wir zeigen, dass immer wenn die Aussage für ein beliebiges $n$ gilt, gilt sie auch für $n+1$: Angenommen, die Aussage gilt für ein beliebiges $n$.
Sei $S(n)$ die Summe der ersten $n$ natürlichen Zahlen, dann folgt aus der Annahme $S(n) = \frac{n\cdot(n+1)}{2}$.
Die Summe der ersten $n+1$ natürlichen Zahlen ist $S(n) + (n+1)$. Es folgt:

$$\begin{aligned} S(n) + (n+1) &= \frac{n\cdot(n+1)}{2} + (n+1) = \frac{n\cdot(n+1) + 2\cdot(n+1)}{2} \\ &= \frac{(n+1)\cdot((n+1)+1)}{2} \end{aligned}$$

Also gilt die Aussage für $(n+1)$ immer dann, wenn sie für $n$ gilt. Da sie für $n = 1$ gilt, folgt aus der Definition der natürlichen Zahlen und dem Induktionsschluss, dass sie für alle natürlichen Zahlen gelten muss.



# Beweis durch vollständige Induktion

## Satz

*Die Summe der ersten $n$ ungeraden Zahlen ist $n^2$. Es gilt also:*

$$\sum_{i=1}^{n} (2i-1) = n^2$$

**Induktionsanfang.** Für $n = 1$ ist die Summe der ersten $n$ ungeraden Zahlen 1. Da $1^2 = 1$ gilt die Aussage für $n = 1$.

**Induktionsschluss.** Wir zeigen, dass immer wenn die Aussage für ein beliebiges $n$ gilt, gilt sie auch für $n+1$: Angenommen, die Aussage gilt für ein beliebiges $n$.
Sei $U(n)$ die Summe der ersten $n$ ungeraden Zahlen, dann folgt aus der Annahme $U(n) = n^2$.
Die Summe der ersten $n+1$ ungeraden Zahlen ist $U(n) + (2(n+1) - 1)$, da $2(n+1) - 1$ die $(n+1)$-te ungerade Zahl ist. Es folgt:

$$U(n)+(2(n+1)-1) = n^2+2(n+1)-1 = n^2+2n+2-1 = n^2+2n+1 = (n+1)^2$$

Also gilt die Aussage für $(n+1)$ immer dann, wenn sie für $n$ gilt. Da sie für $n = 1$ gilt, folgt aus der Definition der natürlichen Zahlen und dem Induktionsschluss, dass sie für alle natürlichen Zahlen gelten muss.



# Beweis durch vollständige Induktion

## Satz

*Es gibt $\binom{n}{k}$ Möglichkeiten aus einer Menge von $n$ Elementen eine Teilmenge mit $k$ Elementen zu bilden, wenn $n \geq k$ gilt.*

Sei $n$ beliebig. Induktion über $k$:

**Induktionsanfang.** Die Aussage gilt für $k = 0$: Es gibt genau eine Teilmenge mit 0 Elementen nämlich die leere Menge. Ferner gilt $\binom{n}{0} = 1$ für beliebiges $n$.

**Induktionsschluss.** Wir zeigen, dass immer wenn die Aussage für ein beliebiges $k$ gilt, gilt sie auch für $k+1$, solange $k+1 \leq n$: Angenommen, die Aussage gilt für ein beliebiges $k$ mit $k+1 \leq n$.
Jede der $k$-elementigen Teilmengen muss um ein Element vergrößert werden. Für jede dieser Mengen gibt es $n-k$ Elemente der Grundmenge, die noch nicht Element der Menge sind und daher zur Vergrößerung hinzugenommen werden können.
Allerdings kann jede der neuen, vergrößerten Teilmengen mit $k+1$ Elementen insgesamt auf $k+1$ Arten aus einer $k$-elementigen Teilmenge durch Hinzunahme eines weiteren Elements entstanden sein.

Es gibt also $\binom{n}{k}\cdot\frac{n-k}{k+1}$ $k$-elementige Teilmengen zu einer beliebigen $n$-elementigen Menge.

$$\binom{n}{k}\cdot\frac{n-k}{k+1} = \frac{n!\cdot(n-k)}{k!\cdot(n-k)!\cdot(k+1)} = \frac{n!}{(k+1)!\cdot(n-k-1)!} = \frac{n!}{(k+1)!\cdot(n-(k+1))!} = \binom{n}{k+1}$$

# Beweis durch vollständige Induktion

## Satz

*In einer Herde mit n Pferden besitzen alle Pferde die gleiche Färbung.*

Induktionsanfang. In einer Herde mit $n = 1$ Pferden haben alle Pferde die gleiche Färbung.

Induktiosschluss. Sei $M$ eine Herde mit $n + 1$ Pferden. Nimmt man ein Pferd $p$ aus $M$ heraus, besitzen die $n$ Pferde in $M \setminus \{p\}$ nach Induktionssannahme die gleiche Färbung $F$. Nimmt man ein weiteres Pferd $q \neq p$ aus $M \setminus \{p\}$ heraus, haben die $n - 1$ Pferde in $M \setminus \{p, q\}$ die Färbung $F$. Nimmt man $p$ wieder hinzu, haben die $n$ Pferde in $M \setminus \{q\}$ die Färbung $F$, insbesondere hat auch $p$ die Färbung $F$. Damit haben alle Pferde in $M$ die Färbung $F$, also die gleiche Färbung.

Was ist schiefgelaufen?Betrachte den Fall $n = 2$

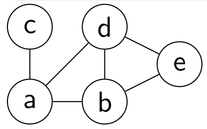


# Eulerscher Polyedersatz

## Theorem

Sei $G$ die planare Zeichnung eines zusammenhängenden, planaren Graphen mit E Ecken, K Kanten, der die Fläche in F Gebiete unterteilt. Dann gilt

$$E - K + F = 2.$$




E=5, K=6, F=3

E-K+F=5-6+3=2



# Beweis durch strukturelle Induktion / Eulerscher Polyedersatz

Der *einfachste* zusammenhängende, planare Graph besteht aus nur einer Ecke. Es gibt eine Fläche (die Außenfläche) und keine Kanten. Es gilt also: $E - K + F = 1 - 0 + 1 = 2$.
In Worten: Knotenzahl - Kantenzahl + Gebietszahl = 2.



Abbildung: Der *einfachste* Graph

Aus diesem *einfachsten* Graphen können alle weiteren ausschließlich durch die beiden folgenden Operationen konstruiert werden (dabei bleibt die Gültigkeit des Satzes erhalten):



# Eulerscher Polyedersatz

1. *Hinzufügen einer Ecke*, die über eine neue Kante mit dem Rest des Graphen verbunden ist. Die Anzahl der Ecken und Kanten steigt jeweils um eins, während die Anzahl der Flächen gleichbleibt. Galt für den alten Graphen $E - K + F = 2$, so gilt es auch für den neuen, da auf der linken Seite der Gleichung je eine Eins addiert und abgezogen wurde.

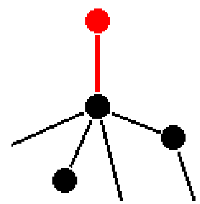


Abbildung: *Hinzufügen einer Ecke*

# Eulerscher Polyedersatz

2. *Hinzufügen einer Kante*, die zwei bereits bestehende Ecken verbindet. Während die Anzahl der Ecken gleichbleibt, steigt die Anzahl der Flächen und Kanten jeweils um eins. Wieder bleibt die Summe $E - K + F$ gleich, da je eine Eins addiert und abgezogen wurde.

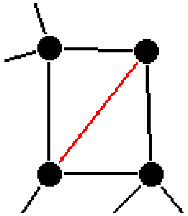


Abbildung: *Hinzufügen einer Kante*

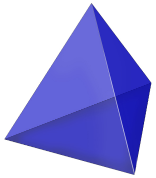


# Eulerscher Polyedersatz

Da der Satz für den ersten, einfachsten Graphen galt, muss er also auch für jeden Graphen gelten, der durch eine der beiden Operationen aus diesem entsteht. Jeder Graph, der durch eine weitere Operation aus einem solchen Graphen entsteht, muss den Satz ebenfalls erfüllen usw.
Daher gilt der Satz für alle zusammenhängenden, planaren Graphen und damit auch für alle konvexen Polyeder.
Der Beweis stammt von Cauchy. Er war der Erste, der das Problem der konvexen Polyeder auf ein Problem der Graphentheorie reduzierte.



# Eulerscher Polyedersatz



**Polyeder** Ein Polyeder ist eine Teilmenge des 3-dimensionalen Raums, welche nur von ebenen Flächen begrenzt wird.

**konvexe Polyeder** Ein Polyeder ist konvex, wenn alle Punkte der Verbindungsstrecke zweier beliebiger Punkte des Polyeders zu dem Polyeder gehören.

### Theorem

Eulerscher Polyedersatz Für ein beschränktes konvexes Polyeder mit $E$ Ecken, $K$ Kanten und $F$ Flächen gilt:

$$E - K + F = 2$$