

1 Mengen

1.1 Grundprinzipien der (naiven) Mengenlehre

Eine *Menge* ist die Zusammenfassung beliebig vieler, beliebiger Objekte zu einem abstrakten Ganzen. Altkanzler Schröder, die Zahl 17 und der Eiffelturm, insgesamt also drei Objekte, lassen sich zum Beispiel zu einer Menge M_1 zusammenfassen; ebenso lassen sich alle durch 5 teilbaren geraden Zahlen, insgesamt also unendlich viele Objekte, zu einer Menge M_2 zusammenfassen; ebenso lässt sich gar nichts, insgesamt also null Objekte, zu einer Menge M_3 zusammenfassen.

Die Objekte, die zu einer Menge M zusammengefasst werden, nennt man die *Elemente von M* . *Elementschaft* ist also ein relativer Begriff, eine Beziehung, die *zwischen* einem Objekt und einer Menge bestehen kann. Um auszudrücken, dass ein Objekt x Element einer Menge M ist, schreibt man ' $x \in M$ '; um das Gegenteil auszudrücken, dass also die Elementschaftsbeziehung zwischen x und M nicht besteht, schreibt man: ' $x \notin M$ '. Es gilt also: $17 \in M_1$, $30 \in M_2$, Schröder $\notin M_3$, etc.

Im allgemeinen handelt es sich bei den Elementen einer Menge um genau diejenigen Objekte, die einer bestimmten Bedingung genügen. Die Elemente von M_2 sind z.B. gerade diejenigen Objekte, die der Bedingung *ist eine gerade Zahl und ist durch 5 teilbar* genügen. Es wird sich im folgenden als geschickt erweisen, Bedingungen stets in der folgenden Form anzugeben:

(*) *so ein x zu sein, dass gilt: ...*

Die eben genannte, für M_2 einschlägige Bedingung lautet dann: *so ein x zu sein, dass gilt: x ist eine gerade Zahl, und x ist durch 5 teilbar*. Die Menge aller x , die einer Bedingung der Form (*) genügen, wird dann durch folgende Notation angegeben:

(+) $\{x \mid \dots\}$

Nach dieser Konvention gilt also:

$$M_2 = \{x \mid x \text{ ist eine gerade Zahl, und } x \text{ ist durch } 5 \text{ teilbar}\}$$

Das Gleichheitszeichen '=' drückt dabei die Identität aus; $a = b$ heißt also, dass es sich bei a und b um dasselbe Objekt handelt.

Auch die Kraut-und-Rüben-Menge M_1 lässt sich in der Form (+) angeben; denn offensichtlich ist ein beliebiges Objekt Element dieser Menge, wenn es entweder mit Gerhard Schröder, mit der Zahl 17 oder mit dem Eiffelturm identisch ist:

$$M_1 = \{x \mid x \text{ ist Gerhard Schröder oder } x \text{ ist } 17 \text{ oder } x \text{ ist der Eiffelturm}\}$$

Um auch M_3 in die Form (+) zu bringen, braucht man irgendeine Bedingung, der überhaupt kein Objekt genügt. Da jedes Ding mit sich selbst identisch ist, wird z.B. die Bedingung *so ein x zu sein, dass gilt: x ist verschieden von x* von keinem Objekt erfüllt:

$$M_3 = \{x \mid x \neq x\}$$

Das Symbol ‘ \neq ’ steht dabei für *Verschiedenheit*, im Sinne von Nicht-Identität. Es ist zu beachten, dass es sich bei (+) um eine Notationskonvention handelt, also eine Verabredung darüber, wie bestimmte Formeln – in diesem Falle die der Form (+) – zu verstehen sind – nämlich als *Bezeichnungen* für entsprechende Mengen: die Notation ‘ $\{x \mid x \text{ ist verschieden von } x\}$ ’ bezeichnet die Menge M_3 – ebenso wie die Bezeichnung ‘ M_3 ’. (Man beachte die Verwendung der Anführungszeichen!)

Eine Menge bestimmt sich einzig und allein durch die Dinge, die in ihr zusammengefasst werden, also ihre Elemente; d.h. es gilt das:

Extensionalitätsprinzip

Wenn eine Menge A genau dieselben Elemente besitzt wie eine Menge B , dann ist $A = B$.

Das Extensionalitätsprinzip kann man weder beweisen noch widerlegen. Es handelt sich vielmehr um eine begriffsbildende Annahme: der Mengenbegriff, also der Begriff der Zusammenfassung von Objekten zu einem abstrakten Ganzen, ist so zu verstehen, dass er dieses Prinzip erfüllt. Das Extensionalitätsprinzip klingt banal, hat aber einige wichtige Konsequenzen, z.B. für die Menge M_3 . Diese Menge hat 0 Elemente; sie ist, wie man sagt, *leer*. Eine unmittelbare Folgerung aus dem Extensionalitätsprinzip ist nun, dass es nur eine solche leere Menge geben kann. Wenn nämlich A und B Mengen sind, die beide kein Element besitzen, dann besitzen sie insbesondere genau dieselben Elemente, womit nach dem Extensionalitätsprinzip gilt: $A = B$. Statt von *einer* leeren Menge darf man also getrost von *der* leeren Menge sprechen. In der Mengenlehre spielt die leere Menge eine wichtige Rolle; sie wird durch das Symbol ‘ \emptyset ’ bezeichnet. Es gilt also: $M_3 = \emptyset$.

Die durch Formeln der Form (+) bezeichneten Mengen fassen jeweils diejenigen Objekte zusammen, die einer bestimmten Bedingung genügen. Die Notation macht natürlich nur dann Sinn, wenn es eine solche Menge überhaupt gibt, wenn sich also alle Objekte, die der betreffenden Bedingung genügen, überhaupt zu einer Menge zusammenfassen *lassen*, wovon wir stets ausgehen werden. Wir unterstellen damit das sogenannte

Komprehensionsprinzip

Für jede Bedingung gibt es eine Menge, die genau diejenigen Objekte zusammenfasst, die dieser Bedingung genügen.

Braucht man überhaupt ein solches Prinzip? Ist es nicht klar, dass jedermann auf ganz willkürliche Weise Objekte zu abstrakten Ganzen zusammenfassen kann, wie wir das eben im Falle von M_1 – M_3 getan haben? Erstaunlicherweise ist dies nicht so klar, wie es zunächst scheint. Denn bei unkontrollierter Verwendung des Komprehensionsprinzips kann man sich sogar in Widersprüche verstricken (s.u.)! Die Entdeckung dieses Sachverhalts (durch B. Russell Anfang des 20. Jahrhunderts) hat die damalige Mathematik in eine tiefe Krise gestürzt und zur Aufgabe des Komprehensionsprinzips geführt; denn wenn in einer Theorie ein Widerspruch steckt, ist sie

wertlos, weil sich in ihr jede Aussage beweisen lässt, also auch jede falsche Aussage. Wir werden uns in diesem Kurs dennoch dieses Prinzips bedienen, um die wichtigsten mengentheoretischen Konstruktionen durchzuführen. Dieses Vorgehen, sich auf vorsichtige Weise des bekanntermaßen widersprüchlichen Komprehensionsprinzips zu bedienen, wird als *naive* Mengenlehre bezeichnet.

Die einfachste Art, aus dem Komprehensionsprinzip einen Widerspruch herzuleiten, ist die sog. *Russellsche Antinomie*, bei der man die “Russellsche” Bedingung (R) *so ein x zu sein, dass x nicht Element von x ist* betrachtet. Nach dem Komprehensionsprinzip gibt es eine Menge aller Objekte, die der Bedingung (R) genügen; nennen wir diese Menge einfach R . R ist offenbar sehr groß, denn sehr viele Objekte erfüllen (R), angefangen mit all denjenigen, die keine Mengen sind (weil sie sowieso keine Elemente haben, also auch nicht sich selbst); aber auch die Mengen $M_1 - M_3$ enthalten sich nicht selbst als Element und genügen damit der Bedingung R . Es gilt also: $M_1 \notin M_1, M_2 \notin M_2, M_3 \notin M_3$, und daher: $M_1 \in R, M_2 \in R, M_3 \in R$. Andererseits gibt es durchaus Objekte, die (R) nicht erfüllen – z.B. die Menge *aller* Mengen, also die Zusammenfassung U aller Objekte, die der Bedingung (M) genügen, *so ein x zu sein, dass gilt: x ist eine Menge*; denn U genügt selbst der Bedingung (M) und ist demnach ein Element von U : $U \in U$. Damit genügt U nicht der Bedingung (R), und somit gilt: $U \notin R$. Genügt R selbst der Bedingung (R)? Gilt also: $R \in R$? Das kann nicht sein, denn in R sind nur solche Objekte, die nicht Element von sich selbst sind. Also muss gelten: $R \notin R$. Aber das kann auch nicht sein, denn dann würde R ja gerade der Bedingung (R) genügen! R kann also weder ein Element noch kein Element von R sein – ein glatter Widerspruch!

Extensionalitäts- und Komprehensionsprinzip werden manchmal zu einem einzigen Prinzip zusammgezogen – dem Extensionalen Komprehensionsprinzip, nach dem es für jede Bedingung *genau eine* Menge gibt, die die Objekte zusammenfasst, die der Bedingung genügen. Aus dem Zusammenspiel von Extensionalitäts- und Komprehensionsprinzip ergibt sich eine wichtige Konsequenz: wenn es bei einer Menge nur darauf ankommt, welche Elemente sie besitzt, dann spielen viele Details in der Formulierung der Elementschaftsbedingung keine Rolle – es kommt nur darauf an, welche Objekte die Bedingung erfüllen. Zum Beispiel erfüllen jeweils genau dieselben Objekte die folgenden beiden Bedingungen:

Bedingung 1:

so ein x zu sein, dass gilt: x ist Gerhard Schröder oder x ist 17 oder x ist der Eiffelturm

Bedingung 2:

so ein x zu sein, dass gilt: x ist die einzige Primzahl zwischen 14 und 18 oder x ist der bislang letzte männliche BRD-Kanzler oder x ist das Wahrzeichen der Pariser Weltausstellung im Jahre 1900 oder x ist die Summe von 13 und 4

Jedes Objekt, das die erste dieser beiden Bedingungen erfüllt, erfüllt auch die zweite, und umgekehrt. Damit besitzt die Menge M_4 der Objekte, die die zweite Bedingung erfüllen, genau dieselben Elemente wie die Menge M_1 . Nach dem Extensionalitätsaxiom gilt dann: $M_4 = M_1$. Natürlich ist es etwas umständlich, wenn man diese Menge als $\{x \mid x \text{ ist die einzige Primzahl zwischen 14 und 18 oder } x \text{ ist der bislang letzte männliche BRD-Kanzler oder } x \text{ ist das Wahrzeichen der Pariser Weltausstellung im Jahre 1900 oder } x \text{ ist die Summe von 13 und 4}\}$ charakterisiert. Aber dieser Unterschied in der Charakterisierung entspricht keinem Unterschied in der charak-

terisierten Sache. Das folgt aus dem Extensionalitätsprinzip.

AUFGABE

Geben Sie jeweils zwei weitere Charakterisierungen der Mengen M_2 und M_3 .

Sehr kleine Mengen wie M_1 werden oft statt durch eine Bedingung durch Auflistung ihrer Elemente charakterisiert:

$$M_1 = \{17, \text{Gerhard Schröder, Eiffelturm}\}$$

Diese *Listennotation* ist natürlich nur praktikabel, wenn es sich um sehr wenige Elemente handelt und wenn diese Elemente einigermaßen knapp bezeichnet werden können. Man muss dabei aufpassen, dass diese Bezeichnungen kein Komma enthalten, weil dieses sonst als Trennstrich zwischen den Elementen gelesen werden kann; die Menge, die aus den Zahlen 0,5 und 1,7 besteht, sollte man also nicht als $\{0,5,1,7\}$ notieren!

Die Listennotation bringt gegenüber der Charakterisierung durch Bedingungen nichts wirklich Neues: Listen lassen sich als Abkürzungen für Bedingungen auffassen. Bezeichnungen der Form

$$\{a_1, \dots, a_n\}$$

stehen dann für entsprechende Charakterisierungen der Form

$$\{x \mid x = a_1 \text{ oder } \dots \text{ oder } x = a_n\}$$

Damit ist auch klar, dass es bei der Listennotation weder auf die Reihenfolge der Elemente noch auf ihre Häufigkeit noch auf ihre genaue Benennung ankommen kann. Zum Beispiel gilt:

$$\begin{aligned} & \{2 + 2, 4^2 + 1, 2 \cdot 2, 20 - 3\} \\ = & \{x \mid x = 2 + 2 \text{ oder } x = 4^2 + 1 \text{ oder } x = 20 - 3 \text{ oder } x = 2 \cdot 2\} \\ = & \{x \mid x = 4 \text{ oder } x = 17 \text{ oder } x = 17 \text{ oder } x = 4\} \\ = & \{x \mid x = 17 \text{ oder } x = 4\} \\ = & \{17, 4\} \end{aligned}$$

Listen können beliebig lang sein – oder kurz. Der Grenzfall sind Listen mit nur einem Eintrag; sie charakterisieren *Einermengen*, die nur ein einziges Element enthalten. Zum Beispiel ist $\{17\}$ die Menge der Objekte, die der Bedingung *so ein x zu sein, dass gilt: $x = 17$* genügen. Natürlich

gibt es nur ein einziges Objekt dieser Art, nämlich die Zahl 17. Diese Zahl darf man nicht mit der Einermenge $\{17\}$ verwechseln. 17 ist keine Menge und enthält also auch kein Element; $\{17\}$ dagegen ist eine Menge mit genau einem Element. Es gilt insbesondere: $17 \notin 17$, aber $17 \in \{17\}$.

Aus dem Extensionalitätsprinzip ergibt sich eine spezielle Konsequenz für Einermengen:

Behauptung

Für beliebige Objekte a und b gilt: $\{a\} = \{b\}$ gdw. $a = b$.

Die Abkürzung ‘gdw.’ steht dabei für *genau dann, wenn*, was soviel heißt wie *wenn das eine, so das andere*.

Ist die Behauptung wirklich richtig? Ja, denn sie lässt sich beweisen, und zwar so:

Beweis

Um die Richtigkeit der Behauptung nachzuweisen, nehmen wir einmal an, wir hätten irgendwelche Objekte a und b . Wir müssen dann zweierelei zeigen: zuerst (\Rightarrow) zeigen wir, dass $a = b$ ist, wenn $\{a\} = \{b\}$; und dann (\Leftarrow) zeigen wir, dass $\{a\} = \{b\}$ ist, wenn $a = b$.

(\Rightarrow): Es gelte $\{a\} = \{b\}$. Nach dem Extensionalitätsprinzip heißt das, dass jedes Element von $\{a\}$ auch ein Element von $\{b\}$ ist. Da $a \in \{a\}$ – a erfüllt ja die Bedingung, mit a identisch zu sein – ist dann insbesondere auch $a \in \{b\}$, d.h. a ist Element der Menge der Objekte, die mit b identisch sind und erfüllt somit selbst diese Bedingung: $a = b$.

(\Leftarrow): Es gelte $a = b$. Dann wird die Bedingung, mit a identisch zu sein, von genau denselben Objekten erfüllt wie die Bedingung, mit b identisch zu sein. Also ist auch die Menge der Objekte, die die eine Bedingung erfüllen, mit der Menge der Objekte, die die andere Bedingung erfüllen, identisch, d.h.: $\{a\} = \{b\}$.

Mit dem Komprehensions- und dem Extensionalitätsprinzip sind die Grundlagen der (naiven) Mengenlehre gelegt. Alles weitere lässt sich aus diesen beiden Prinzipien herleiten.

1.2 Mengenoperationen

Mengen lassen sich auf vielfältige Weise miteinander kombinieren. In diesem Abschnitt werden wir die elementarsten dieser Kombinationen einführen, *Schnitt*, *Vereinigung* und *Differenz*:

Definitionen

Für beliebige Mengen A und B ist:

- der *Schnitt* von A mit B die Menge derjenigen Objekte x , für die gilt: x ist ein Element von A und x ist ein Element von B ; symbolisch:

$$A \cap B = \{x \mid x \in A \text{ und } x \in B\}$$

- die *Vereinigung* von A und B die Menge derjenigen Objekte x , für die gilt: x ist ein Element von A oder x ist ein Element von B (oder beides); symbolisch:

$$A \cup B = \{x \mid x \in A \text{ oder } x \in B\}$$

- die *Differenz* von A und B die Menge derjenigen Objekte x , für die gilt: x ist ein Element von A , aber x ist kein Element von B ; symbolisch:

$$A \setminus B = \{x \mid x \in A \text{ und } x \notin B\}$$

Die Notation ' $A \cap B$ ' liest man auch als ' A geschnitten (mit) B '; ' $A \cup B$ ' liest man als ' A vereinigt (mit) B '; und ' $A \setminus B$ ' liest man ' A minus B ' oder ' A ohne B '.

Hier sind ein paar Beispiele (Stadt-Land-Fluss):

- für Schnitte:

$$\{\text{Frankfurt, Kassel, Stuttgart}\} \cap \{\text{Kassel, Frankfurt, Berlin}\} = \{\text{Kassel, Frankfurt}\}$$

$$\{x \mid x \text{ hat mehr als } 500\,000 \text{ Einwohner}\} \cap \{x \mid x \text{ liegt in Hessen}\} = \{\text{Frankfurt}\}$$

$$\{\text{Frankfurt, Kassel, Stuttgart}\} \cap \{x \mid x \text{ liegt in Hessen}\} = \{\text{Kassel, Frankfurt}\}$$

$$\{\text{Frankfurt, Kassel, Stuttgart}\} \cap \{\text{Kassel, Stuttgart}\} = \{\text{Kassel, Stuttgart}\}$$

$$\{x \mid x \text{ ist Landeshauptstadt}\} \cap \{\text{Kassel, Frankfurt}\} = ???$$

- für Vereinigungen:

$$\{\text{Belgien, Niederlande, Luxemburg}\} \cup \{\text{Frankreich, Belgien, Schweiz}\} =$$

$$\{\text{Belgien, Niederlande, Luxemburg, Frankreich, Schweiz}\}$$

$$\{x \mid x \text{ ist Nachbarland von Frankreich}\} \cup \{x \mid x \text{ ist Nachbarland von Belgien}\} =$$

$$\{x \mid x \text{ ist Nachbarland von Frankreich}\} \cup \{\text{Frankreich, Niederlande}\}$$

$$\{\text{Belgien, Niederlande}\} \cup \{\text{Belgien}\} = \{\text{Belgien, Niederlande}\}$$

$$\emptyset \cup \{\text{Frankreich, Niederlande}\} = ???$$

- für Differenzen:

$$\{\text{Rhein, Main}\} \setminus \{\text{Rhein}\} = \{\text{Main}\}$$

$$\{\text{Rhein, Main}\} \setminus \{\text{Elbe}\} = ???$$

$$\{x \mid x \text{ fließt durch Hessen}\} \setminus \{x \mid x \text{ fließt durch Deutschland}\} = ???$$

$$\emptyset \setminus \{\text{Rhein}\} = ???$$

AUFGABE

Ersetzen Sie die ???

Für diese drei, auch als *Boolesche Mengenoperationen* bekannten Kombinationen gelten ein paar einfache Rechenregeln:

Satz 1

Für beliebige Mengen A , B und C gilt:

(IS)	$A \cap A = A$	Idempotenz von \cap
(KS)	$A \cap B = B \cap A$	Kommutativität von \cap
(AS)	$A \cap (B \cap C) = (A \cap B) \cap C$	Assoziativität von \cap
(LS)	$A \cap \emptyset = \emptyset$	
(IV)	$A \cup A = A$	Idempotenz von \cup
(KV)	$A \cup B = B \cup A$	Kommutativität von \cup
(AV)	$A \cup (B \cup C) = (A \cup B) \cup C$	Assoziativität von \cup
(LV)	$A \cup \emptyset = ???$	
(Diff1)	$A \setminus A = ???$	
(Diff2)	$A \setminus \emptyset = ???$	
(Dis1)	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	Distributivgesetz
(Dis2)	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	Distributivgesetz

AUFGABE

Ersetzen Sie die ???

Diese Rechenregeln sind keine neuen Prinzipien, sondern folgen allesamt aus den Definitionen der genannten Operationen. Aufgrund der folgenden elementaren Zusammenhänge lassen sie sich mit Hilfe aussagenlogischer Wahrheitstabeln beweisen:

- (i) $x \in A \cap B$ gdw. $x \in A$ und $x \in B$
- (ii) $x \in A \cup B$ gdw. $x \in A$ oder $x \in B$
- (iii) $x \in A \setminus B$ gdw. $x \in A$ und nicht: $x \in B$

(i)–(iii) folgen unmittelbar aus den Definitionen von \cap , \cup und \setminus ; die hervorgehobenen metasprachlichen Junktoren sind dabei im Sinne der Wahrheitstabeln für \wedge , \vee und \neg zu verstehen. Diesen Sachverhalt kann man für den Nachweis der obigen Gleichungen ausnutzen. So ergibt sich nach (i)–(iii) für die beiden Mengenterme in (Dis1):

- (iv) $x \in (A \cup (B \cap C))$ gdw. $x \in A$ oder: $[x \in B$ und $x \in C]$
- (v) $x \in ((A \cup B) \cap (A \cup C))$ gdw. $[x \in A$ oder $x \in B]$ und: $[x \in A$ oder $x \in C]$

Rechts vom ‘gdw.’ in (iv) und (v) stehen jeweils aussagenlogisch äquivalente Kombinationen von ‘ $x \in A$ ’, ‘ $x \in B$ ’ und ‘ $x \in C$ ’, wie man per Wahrheitstafel leicht nachweist. Da diese

Kombinationen für beliebige x , A , B und C stets zum selben Wahrheitswert führen, sind nach (iv) und (v) auch die Aussagen links vom ‘gdw.’ äquivalent. Nach dem Extensionalitätsaxiom sind dann auch die beiden in (Dis1) gleichgesetzten Mengen in der Tat stets gleich. Für die anderen Gleichungen argumentiert man analog.

Jede der obigen Formeln lässt sich in diesem Stil beweisen. Wir werden uns damit nicht weiter aufhalten, aber wer an dieser Stelle Verständnisprobleme hat, sollte die beiden Musterbeweise im Detail studieren, den Argumentationsgang nachvollziehen und eine der anderen Formeln selbst beweisen. Die Sachverhalte, die diese Formeln ausdrücken, sind relativ durchsichtig. Sie lassen sich auch graphisch einsehen, und zwar anhand so genannter *Venn-Diagramme*, in denen Mengen als Kreise (oder andere Flächen) dargestellt werden:

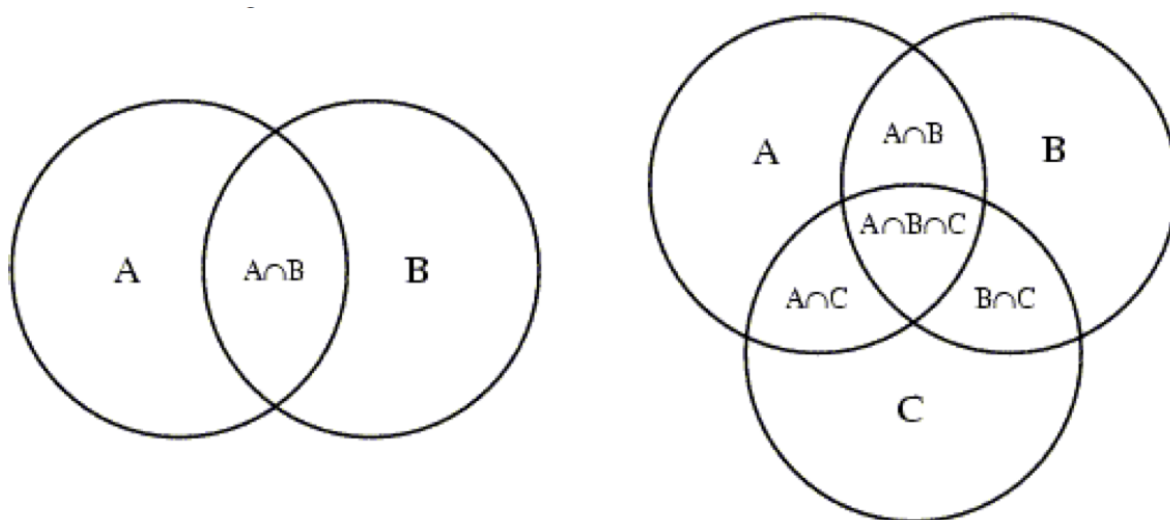


Abbildung 1: Vorsicht: In einem Venn-Diagramm steht jede Region (= Fläche zwischen zwei Begrenzungen) für eine Menge, die allerdings leer sein kann!

1.3 Mengenbeziehungen

Mengen können in verschiedenen Beziehungen zueinander stehen. Betrachten wir dazu ein paar Beispiele:

$$A = \{1, 2, 3, 4\}; B = \{2, 4\}; C = \{3, 4, 5\}; D = \{6, 7\}$$

Die letzte Menge, D , fällt insofern aus dem Rahmen, als keines ihrer Elemente in einer der drei anderen Mengen zu finden ist; D ist von jeder drei anderen Mengen *disjunkt*. Ansonsten *überlappen* sich die Mengen in dem Sinne, dass jeweils zwei von ihnen (mindestens) ein Element gemeinsam haben. B hat mit A sogar alle seine Elemente gemeinsam; es ist eine *Teilmenge* von A :

Definition

Es seien A und B beliebige Mengen.

A ist disjunkt von B heißt: $A \cap B = \emptyset$.

A überlappt sich mit B heißt: $A \cap B \neq \emptyset$.

A ist eine Teilmenge von B heißt: jedes Element von A ist ein Element von B .

(Man überlege sich, dass diese Definition von Disjunktheit und Überlappung dasselbe besagen wie die informellen Charakterisierungen im Absatz davor!)

Statt ‘ A ist eine Teilmenge von B ’ schreibt man auch kurz: ‘ $A \subseteq B$ ’. Die Teilmengenbeziehung ist für die Mengenlehre ebenso wichtig wie die Elementbeziehung, aber man darf die beiden nicht miteinander verwechseln. Sie haben auch vollkommen verschiedene Eigenschaften. Zum Beispiel gelten die folgenden allgemeinen Gesetze nur für die Teilmengenbeziehung:

Satz 2

Es seien A, B und C beliebige Mengen. Dann gilt:

1. $A \subseteq A$.
2. $\emptyset \subseteq A$.
3. Wenn $A \subseteq B$ und $B \subseteq C$, dann ist $A \subseteq C$.
4. Wenn $A \subseteq B$ und $B \subseteq A$, dann ist $A = B$.

Beweis

ad 1. Jedes Element einer Menge A ist natürlich ein Element von A .

ad 2. Das mag überraschen. Zu zeigen ist aber nur, dass für jedes x gilt: wenn $x \in \emptyset$, dann ist $x \in A$. Hier argumentiert man am besten *per Kontraposition*, also indirekt, indem man zeigt: wenn $x \notin A$, dann ist $x \notin \emptyset$. Sei also $x \notin A$. Dann bleibt zu zeigen, dass $x \notin \emptyset$. Aber das ist sowieso klar, denn \emptyset enthält per definitionem keine Elemente!

ad 3. Angenommen, (i) jedes Element von A ist ein Element von B und (ii) jedes Element von B ist ein Element von C . Wir müssen dann für beliebige $x \in A$ zeigen, dass $x \in C$. Aber nach (i) gilt mit $x \in A$ auch: $x \in B$, woraus nach (ii) folgt: $x \in C$, was zu zeigen war.

ad 4. Das ist das Extensionalitätsprinzip!

AUFGABE

Zeigen Sie anhand von Gegenbeispielen, dass die Gesetze 1. – 3. für die Element-Beziehung nicht immer gelten. Finden Sie also für jede der Aussagen 1’. – 3’. ein Beispiel, das die Aussage widerlegt. (Für verschiedene Aussagen kann man verschiedene Gegenbeispiele nehmen!)

- 1.’ $A \in A$

2.' $\emptyset \in A$

3.' Wenn $A \in B$ und $B \in C$, dann ist $A \in C$.

Lassen sich für 1.' – 3.' auch (positive) Beispiele finden?

Das allgemeine Gesetz 1. zeigt, dass man mit dem Teilmengenbegriff vorsichtig umgehen muss: im Gegensatz zum alltäglichen Sprachgebrauch ist in der Mengenlehre jede Menge Teil von sich selbst. Die Aussage, dass $A \subseteq B$, schließt also im allgemeinen nicht aus, dass $A = B$. Will man dagegen zum Ausdruck bringen, dass jedes Element einer Menge A auch Element einer Menge B ist, aber nicht umgekehrt, muss man einen anderen Begriff als den der Teilmenge wählen:

Definition

A ist eine echte Teilmenge von B – symbolisch: $A \subsetneq B$ – heißt: $B \neq A \subseteq B$.

Wir haben hier die Beziehungen ‘durchgeschrieben’: $B \neq A \subseteq B$ ist zu verstehen im Sinne der doppelten Bedingung: $B \neq A$ und $A \subseteq B$.

1.4 Potenzmengen

Die Teilmengen einer Menge A erfüllen alle die Bedingung *so ein x zu sein, dass gilt: $x \subseteq A$* . Nach dem Komprehensionsprinzip lassen sie sich also wieder zu einer Menge zusammenfassen:

Definition

Sei A eine beliebige Menge. Die *Potenzmenge* von A – symbolisch: $\wp(A)$ – ist die Menge $\{x \mid x \subseteq A\}$.

Wie sieht die Potenzmenge einer Menge A im allgemeinen aus? Zunächst enthält sie für jedes Element $x \in A$ auch dessen Einermenge $\{x\}$ und außerdem noch – nach Punkt 2. von Satz 2 – die leere Menge. $\wp(A)$ enthält also mindestens ein Element mehr als A ; wir werden noch sehen, dass $\wp(A)$ im allgemeinen sehr viel größer ist als A . Anhand eines konkreten Beispiels lässt sich das aber schon jetzt erahnen. Die Teilmengen der drei-elementigen Menge $\{0, 5, 8\}$ etwa sind – geordnet nach ihrer Größe:

0 Elemente	\emptyset	1
1 Element	$\{0\}, \{5\}, \{8\}$	3
2 Elemente	$\{0, 5\}, \{0, 8\}, \{5, 8\}$	3
3 Elemente	$\{0, 5, 8\}$	1

(Die rechte Spalte gibt die Anzahl der Mengen in der mittleren Spalte an.) Insgesamt handelt es sich also um 8 Teilmengen, d.h. $\wp(\{0, 5, 8\})$ enthält 8 Elemente:

$$\wp(\{0, 5, 8\}) = \{\emptyset, \{0\}, \{5\}, \{8\}, \{0, 5\}, \{0, 8\}, \{5, 8\}, \{0, 5, 8\}\}$$

Die Teilmengen einer Menge A – also die Elemente von $\wp(A)$ – können untereinander wieder in der Teilmengenbeziehung stehen. Wie man an der obigen Tabelle leicht nachprüft, ist jede Menge in den oberen drei Reihen Teilmenge mindestens einer Menge in der Reihe unter ihr. Die folgende Graphik – ein sog. *Hasse-Diagramm* – veranschaulicht sämtliche Teilmengenbeziehungen in $\wp(\{0, 5, 8\})$:

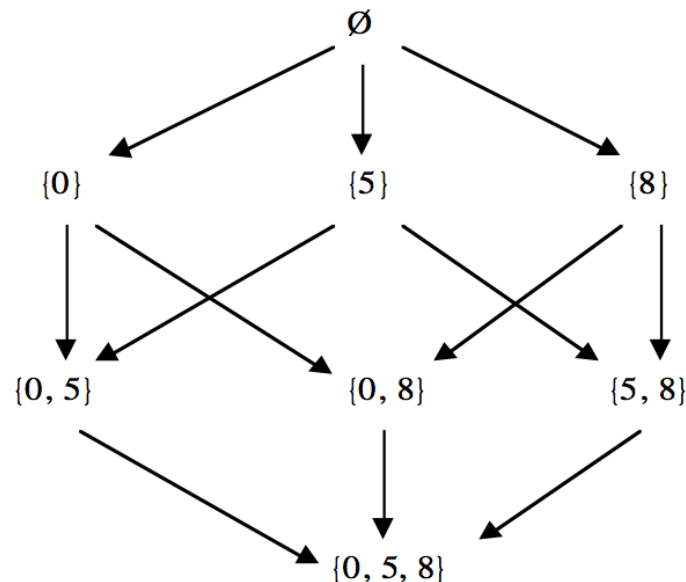


Abbildung 2: Hasse-Diagramm einer Potenzmenge

Die Graphik erklärt sich praktisch von selbst: $A \subseteq B$ gilt gerade, wenn es eine (direkte oder indirekte) Pfeilverbindung gibt, die bei A anfängt und mit B aufhört.

Ein genauerer Blick auf die Graphik zeigt, dass sich die Vereinigung zweier Teilmengen von $\{0, 5, 8\}$ immer an der Stelle befindet, die von den beiden Mengen aus mit der geringsten Anzahl von Pfeilen erreichbar ist; das gleiche gilt für den Schnitt – nur dass man in dem Fall die Pfeile *rückwärts* verfolgen muss. Die graphische Darstellung der Potenzmenge von $\{0, 5, 8\}$ birgt also mehr Struktur, als man auf den ersten Blick sieht. Jede Potenzmenge lässt sich im Prinzip mit einem solchen Pfeildiagramm darstellen, das oben bei der leeren Menge beginnt, zur Mitte hin breiter wird, dann wieder abnimmt und schließlich bei der Gesamtmenge endet. Es ist oft zweckmäßig, sich diese Zwiebelstruktur vor Augen (!) zu halten, wenn von einer Potenzmenge die Rede ist.

Die im vorhergehenden Abschnitt betrachteten Mengenoperationen lassen sich immer innerhalb einer Potenzmenge durchführen: der Schnitt, die Vereinigung und die Differenz zweier Teilmengen einer Menge U sind stets auch wieder in $\wp(U)$. (WARUM???) Für das Rechnen mit Mengenoperationen *innerhalb einer Potenzmenge* gelten – neben den in Satz 1 aufgelisteten –

einige spezielle Rechenregeln, die auf die Gesamtmenge U Bezug nehmen:

Satz 3

Es sei U eine beliebige Menge. Dann gilt für alle Teilmengen A und B von U :

	$\overline{\emptyset} = U$	
	$\overline{U} = \emptyset$	
(N)	$(A \cap U) = A$	Neutralität von U
(\hat{N})	$(A \cup U) = U$	Neutralität von U
(TND)	$(A \cup \overline{A}) = U$	Tertium non datur
(W)	$(A \cap \overline{A}) = \emptyset$	Satz vom Widerspruch
(DM1)	$\overline{(A \cup B)} = (\overline{A} \cap \overline{B})$	de Morgansches Gesetz
(DM2)	$\overline{(A \cap B)} = (\overline{A} \cup \overline{B})$	de Morgansches Gesetz

Die Überstreichung bezeichnet dabei das so genannte *Komplement*, das ist die Differenz relativ zum ‘Universum’ U : $\overline{A} = (U \setminus A)$. Diese Notation ist hochgradig kontextabhängig (und deswegen leicht irreführend), denn sie nimmt implizit auf eine vorgegebene Menge U Bezug, ohne sie explizit zu nennen. Man sollte sie also nur verwenden, wenn vorher klar und deutlich gesagt wird, was diese Menge U (das betrachtete ‘Universum’) genau ist. Ohne eine solche Angabe ist eine Bezeichnung wie ‘ $\{0, 1, 5\}$ ’ nicht eindeutig!

Die in Satz 3 genannten Gleichungen lassen sich wieder mit Wahrheitstabellen nachweisen oder mit Hilfe von Venn-Diagrammen einsehen. Dabei ist zu beachten, dass für Teilmengen A und Elemente x des Universums U die Bedingungen $x \in \overline{A}$ und $\neg x \in A$ (d.h. $x \notin A$) äquivalent sind. In Letzteren stellen wir das Universum U als Rechteck dar, in dem die Teilmengen A und B liegen. Der Schnitt von A und B sieht dann so aus:

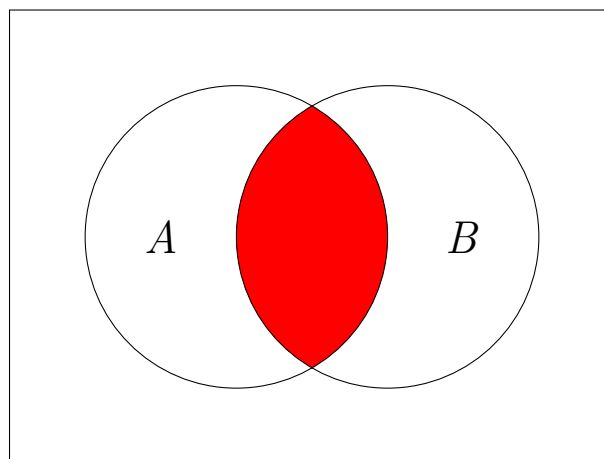


Abbildung 3: $A \cap B$

Auf der linken Seite von (DM2) steht das Komplement dieses Schnittes (relativ zu U), also die Menge aller Elemente von U , die nicht in diesem Schnitt sind:

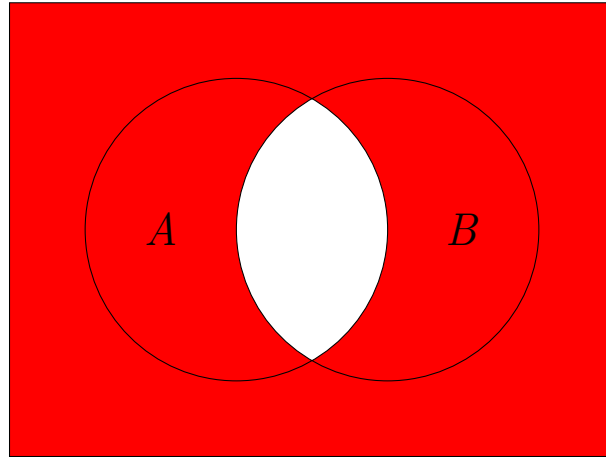


Abbildung 4: $\overline{A \cap B}$

Um zu sehen, wie die Menge auf der rechten Seite aussieht, betrachtet man zunächst die beiden Komplemente, aus denen sie gebildet ist, also:

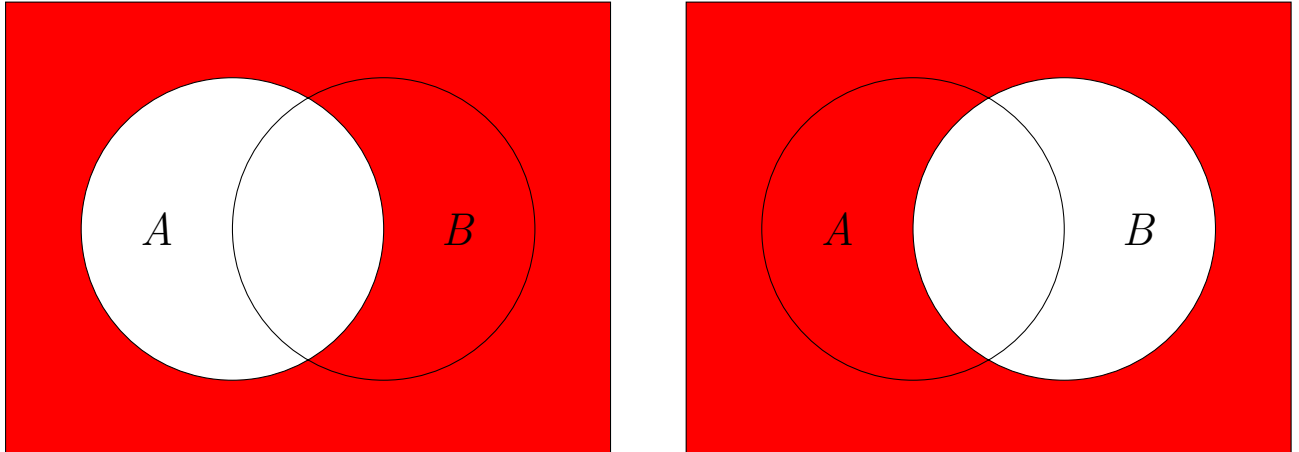


Abbildung 5: \overline{A} bzw. \overline{B}

Die Vereinigung dieser beiden Mengen deckt alles ab, was in einer von beiden liegt, also die Regionen, die links oder rechts (oder links und rechts) markiert sind – also alles außer der Schnittregion in der Mitte. Und das ist dieselbe Menge wie das oben dargestellte $\overline{A \cap B}$.

AUFGABE

Beweisen Sie Satz 3 mit Hilfe von Wahrheitstafeln, und illustrieren Sie (DM1) mit Venn-Diagrammen.

Wir beenden unsere Betrachtungen zu Potenzmengen mit einem Spezialfall, der leeren Menge. Wie sieht ihre Potenzmenge aus? Was sind die Teilmengen von \emptyset ? Man könnte meinen, gar keine, aber das ist falsch. Denn laut Satz 1 besitzt jede Menge schon einmal sich selbst als Teilmenge; eine der Teilmengen von \emptyset ist also \emptyset selbst. (Ebenso ist \emptyset Teilmenge jeder Menge, aber das bringt hier nichts Neues.) Hat \emptyset sonst noch irgendwelche Teilmengen? Es sieht nicht so aus, aber können wir das auch beweisen? Wir können. Wäre $A \subseteq \emptyset$, aber $A \neq \emptyset$, dann müsste A ein Element enthalten; denn sonst hätten ja A und \emptyset dieselben Elemente (Extensionalitätsprinzip!). Wäre aber $a \in A$, dann wäre auch $a \in \emptyset$ (weil $A \subseteq \emptyset$) was nicht sein kann, weil \emptyset keine Elemente hat. Also kann es nicht angehen, dass $A \subseteq \emptyset$, aber $A \neq \emptyset$. M.a.W.: \emptyset ist die einzige Teilmenge von \emptyset : $\wp(\emptyset) = \{\emptyset\}$. Man beachte, dass das nicht heißt, dass die leere Menge gleich ihrer Potenzmenge ist. Denn $\{\emptyset\}$ ist nicht leer, sondern enthält vielmehr ein Element, nämlich \emptyset ; $\wp(\emptyset)$ ist also eine Einermenge.

AUFGABE

Wie viele Elemente haben die Potenzmengen von $\{0\}$ und $\{0, 1\}$? Welche? Stellen Sie $\wp(\{0\})$ und $\wp(\{0, 1\})$ durch Pfeildiagramme dar.

2 Relationen

2.1 Tupel

Das Extensionalitätsprinzip besagt, dass es bei einer Menge nur auf die Elemente ankommt. Insbesondere kommt es bei der Listendarstellung einer Menge weder auf die Reihenfolge noch auf die Häufigkeit der Nennung eines Elements an. Manchmal ist es jedoch zweckmäßig, einzelne Objekte auf geordnete Weise zusammenzufassen, so dass gerade Reihenfolge und Häufigkeit eine Rolle spielen.

Das gilt zum Beispiel für die Darstellung einfacher mathematischer Strukturen wie dem kartesischen Koordinatenkreuz. Es liegt nahe, das gesamte Kreuz als Menge aller Punkte (x, y) aufzufassen, wobei x und y reelle Zahlen sind. Aber was ist ein Punkt wie $(-1, +1)$? Man darf ihn jedenfalls nicht mit der Menge $\{-1, +1\}$ identifizieren, denn dann ginge der Unterschied zum spiegelbildlichen Punkt $(+1, -1)$ verloren. Ein weiteres Beispiel ist die Anordnung der natürlichen Zahlen nach ihrer Größe und die sich daraus ergebende Beziehung *ist größer als*, symbolisch: ' $>$ '. Diese Beziehung besteht immer zwischen zwei Zahlen. Lässt sich diese Beziehung mengentheoretisch darstellen? Man könnte es mit der Menge aller Mengen $\{n, m\}$ versuchen, so dass $n > m$, aber wie wollte man dann die Kleiner-Beziehung – symbolisch ' $<$ ' – modellieren? Und wie würden sich die beiden von der Verschiedenheit – symbolisch ' \neq ' – unterscheiden?

Neben Mengen muss es also noch strukturierte Zusammenfassungen von Objekten geben. Solche strukturierten Zusammenfassungen heißen in der Mengenlehre *Tupel*. Wenn n eine natürliche Zahl ist $(0, 1, 2, \dots)$, hat ein n -Tupel die Gestalt (a_1, \dots, a_n) . Dabei sind a_1, \dots, a_n irgendwelche Objekte. Statt des Extensionalitätsprinzips, das die Identität zwischen Mengen erklärt, gilt für Tupel die folgende Festlegung:

Identitätskriterium für Tupel

Es sei n eine natürliche Zahl, $a_1, \dots, a_n, b_1, \dots, b_n$ seien irgendwelche Objekte. Dann gilt:

$$(a_1, \dots, a_n) = (b_1, \dots, b_n) \text{ gdw. } a_1 = b_1 \text{ und } \dots \text{ und } a_n = b_n.$$

Mit den Tupeln haben wir scheinbar das Begriffsinventar der Mengenlehre – *Menge* und *Element* – erweitert, aber nur scheinbar. Denn im Prinzip lassen sich Tupel aus Mengen fabrizieren – mit der sog. *Kuratowski-Paar-Konstruktion*, nach der das Paar (a_1, a_2) mit der Menge $\{\{a_1, a_2\}, \{a_1\}\}$ und das $n + 1$ -Tupel (a_1, \dots, a_{n+1}) mit dem Paar $(a_1, (a_2, \dots, a_{n+1}))$ identifiziert wird. Wir verzichten hier aus Zeitgründen auf diese Möglichkeit und tun stattdessen so, als handele es sich beim *Tupel* um einen weiteren Grundbegriff.

Man beachte, dass ein Tupel immer endlich lang ist. Z.B. lassen sich nicht alle natürlichen Zahlen in einem Tupel unterbringen. Um unendlich viele Objekte aufzureihen, benötigt man den in Teil 3. eingeführten Begriff der *Folge*.

Um das Kriterium und den Tupelbegriff wirklich auf *alle* natürlichen Zahlen anzuwenden, müssen wir noch festlegen, was 0-Tupel sind. Da für $n = 0$ die Liste der a_1, \dots, a_n keinen Eintrag hat, hat

das einzige 0-Tupel die Gestalt: $()$. Wir werden dieses leere Klammerpaar mit der leeren Menge identifizieren: $() = \emptyset$. 1-Tupel haben dagegen die Gestalt (a) , wobei a jeweils irgendein Objekt ist. Es würde nichts ausmachen, wenn wir in diesem Fall die runden Klammern als Mengenklammern auffassen; praktischer ist es aber (für diverse Zwecke), wenn man die Klammern im Falle von 1-Tupeln ‘überliest’ und für beliebige Objekte a festlegt: $(a) = a$. Man beachte, dass daraus unmittelbar folgt: $((a)) = (a) = a$.

Nach dem obigen Identitätskriterium gilt z.B.: $(1, 1, 0) \neq (0, 1, 1)$, denn die beiden 3-Tupel unterscheiden sich u.a. im ersten Glied. Andererseits sind $(1 + 1, 2)$ und $(2, 1 + 1)$ dasselbe 2-Tupel; denn bekanntlich ist $1 + 1 = 2$. Apropos 2-Tupel und 3-Tupel: in der Mengenlehre bezeichnet man diese auch gern als (geordnete) *Paare* bzw. *Tripel*. 4-Tupel heißen entsprechend *Quadrupel*, 5-Tupel *Quintupel* etc.; aber so lange Tupel werden wir ohnehin nur selten betrachten.

Wie das n -Tupel (a_1, \dots, a_n) von der Menge $\{a_1, \dots, a_n\}$ zu unterscheiden ist, ist auch die Beziehung eines einzelnen a_i zur Menge $\{a_1, \dots, a_n\}$ von der zwischen a_i und dem Tupel (a_1, \dots, a_n) zu unterscheiden. Erstere ist die Elementbeziehung; letztere werden wir als *Komponenten-Beziehung* bezeichnen: a_i ist die i -te Komponente des n -Tupels (a_1, \dots, a_n) . Es gilt also: die zweite Komponente des Paares $(+1, -1)$ ist die Zahl -1 , das Tripel $(0, 0, 0)$ hat drei gleiche Komponenten etc.

Aus Tupeln lassen sich wieder Mengen bilden. Zum Beispiel kann man alle Paare (x, y) von Elementen einer Menge A zu einer neuen Menge A^2 zusammenfassen:

$$A^2 = \{(x, y) \mid x \in A \text{ und } y \in A\}.$$

Die Notation ist dabei so zu verstehen, dass man hier die Bedingung betrachtet, *so ein Paar der Gestalt (x, y) zu sein*, dass gilt: $x \in A$ und $y \in A$.

Wenn A die Menge der reellen Zahlen ist, ist A^2 gerade die Menge der Punkte des schon erwähnten Koordinatenkreuzes. Wegen dieses Zusammenhangs bezeichnet man im allgemeinen (d.h. für beliebige A) die Menge A^2 als das *kartesische Produkt von A (mit sich selbst)*. Hier ist eine noch allgemeinere:

Definition

Es sei n eine natürliche Zahl, A_1, \dots, A_n seien irgendwelche Mengen. Dann ist das *n -stellige kartesische Produkt* von A_1, \dots, A_n wie folgt definiert:

$$A_1 \times \dots \times A_n = \{(x_1, \dots, x_n) \mid x_1 \in A_1 \text{ und } \dots x_n \in A_n\}.$$

Für den Fall, dass $A_1 = \dots = A_n = B$ schreibt man statt ‘ $A_1 \times \dots \times A_n$ ’ auch ‘ B^n ’.

Wir werden es im folgenden vor allem mit dem Fall $n = 2$ und $A_1 = A_2$ zu tun haben, gelegentlich aber auf die allgemeinere Definition zurückgreifen.

Geordnete Paare lassen sich nun auch verwenden, um Beziehungen durch Mengen zu modellie-

ren. Mengentheoretisch gesprochen ist die Kleiner-Beziehung zwischen natürlichen Zahlen die Menge

$$R_{<} = \{(n, m) \mid n \text{ und } m \text{ sind natürliche Zahlen und } n < m\}.$$

AUFGABE

Geben Sie die mengentheoretische Entsprechung $R_{>}$ der Größer-Beziehung zwischen natürlichen Zahlen an und zeigen Sie, dass $R_{<}$ und $R_{>}$ disjunkt sind.

Jedes Element (n, m) der Menge $R_{<}$ ist ein Paar, dessen Komponenten Elemente der Menge \mathbb{N} der natürlichen Zahlen sind. $R_{<}$ ist also eine Teilmenge von \mathbb{N}^2 oder, wie man in der Mengenlehre sagt, eine *zweistellige Relation über \mathbb{N}* . Hier ist die allgemeine:

Definition

Es sei A eine Menge und n eine natürliche Zahl. Eine *n -stellige Relation* (über A) ist eine Teilmenge von A^n .

Wenn R eine zweistellige Relation ist, schreibt man auch ' xRy ' statt ' $(x, y) \in R$ ' und ' $xRyRz$ ' statt ' $(x, y) \in R$ und $(y, z) \in R$ '. Die Menge der Objekte, zwischen denen R besteht, heißt der *Bereich von R* ; der Bereich besteht, genauer gesagt, aus allen $x \in A$, die zu irgendeinem $y \in A$ in der Beziehung R stehen und allen $y \in A$, zu denen irgendein $x \in A$ in der Beziehung R steht. Der Bereich von R ist also eine Teilmenge von A ; aber A kann durchaus mehr Elemente enthalten.

2.2 Die Klassifikation zweistelliger Relationen

Relationen – und vor allem zweistellige – spielen in allen Anwendungen der Mengenlehre eine zentrale Rolle. Es ist üblich, sie ihrer Struktur nach zu klassifizieren. Schauen wir uns dazu erst ein paar Beispiele an:

$$R_\alpha = \{(x, y) \mid y \text{ folgt unmittelbar auf } x \text{ im lateinischen Alphabet}\}.$$

$$R_{\leq} = \{(n, m) \mid n \text{ und } m \text{ sind natürliche Zahlen und } n \leq m\}.$$

$$R_N = \{(n, m) \mid n \text{ und } m \text{ sind natürliche Zahlen und } m = n + 1\}.$$

$$R_A = \{(A, B) \mid A \text{ und } B \text{ sind Personen und } A \text{ ist (an Jahren) jünger als } B\}.$$

$$R_P = \{(M, N) \mid M \subseteq \{0, 5, 8\}, N \subseteq \{0, 5, 8\} \text{ und } M \subseteq N\}.$$

$$R_V = \{(A, B) \mid A \text{ und } B \text{ sind Personen mit demselben Vornamen}\}.$$

In der Definition von R_α werden ' x ' und ' y ' natürlich als Variablen für Buchstaben benutzt. Die Relation lässt sich graphisch sehr einfach durch Pfeile darstellen:

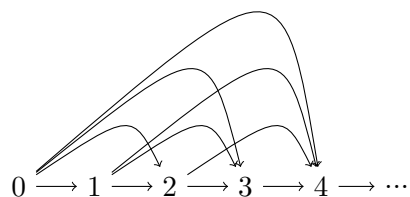
$$A \rightarrow B \rightarrow C \rightarrow \dots \quad \longrightarrow \quad X \rightarrow Y \rightarrow Z$$

Hier besagt ein Pfeil von einem Buchstaben x zum nächsten (y), dass y der alphabetische Nachfolger von x ist, dass also gilt: $xR_\alpha y$. Man beachte, dass in dieser Darstellung von R_α die Relation wirklich nur dort besteht, wo auch ein Pfeil eingezeichnet ist.

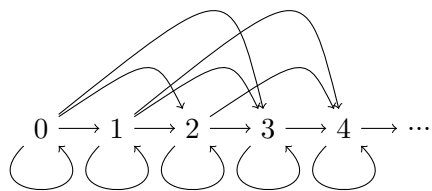
Man kann sich jede zweistellige Relation über einer Menge A als ein Geflecht von Pfeilen vorstellen, die immer dann Elemente x und y von A verbinden, wenn xRy . Allerdings kann man im Falle einer unendlich großen Relation das entsprechende Diagramm nur ausschnittsweise zu Papier bringen. Das gilt z.B. für R_\leq . Man könnte zunächst meinen, dass die Pfeile ähnlich wie bei R_\leq verlaufen, nur unendlich lange, also etwa so:

$$0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4 \longrightarrow \dots$$

Doch halt! Hier haben wir den einen oder anderen Pfeil vergessen. In dieser Darstellung stehen die Pfeile immer nur zwischen einer Zahl und ihrem Nachfolger. Die dargestellte Relation ist also R_N und nicht R_\leq ! Aber R_\leq besteht z.B. auch zwischen 0 und 2, 1 und 4 usw. – also müssen wir auch zwischen diese Zahlen Pfeilverbindungen einbauen:



Aber auch das genügt nicht. Denn die so dargestellte Relation ist $R_<$ – die sich sich ja von R_\leq dadurch unterscheidet, dass jede Zahl zu sich selbst in dieser Relation steht, dass also stets gilt: $nR_\leq n$ (aber nicht $nR_< n$). Wir müssen die Zeichnung also noch um ‘Selbstverweise’ ergänzen:



Sowohl die oberen als auch die unteren Pfeile in dieser Darstellung sind in gewisser Weise vorhersagbar:

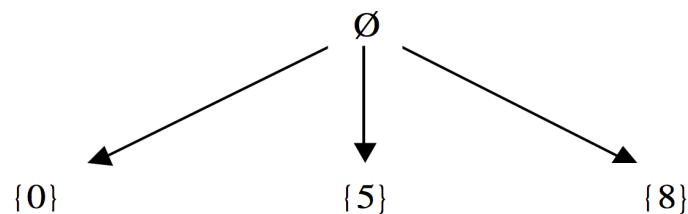
- Die unteren Pfeile ergeben sich, weil *jedes Element* von \mathbb{N} zu sich in der Relation R_\leq steht. Die Relation R_\leq ist – im Gegensatz zu R_α , $R_<$ und R_N – *reflexiv*.

- Die oberen Pfeile ergeben sich aus den mittleren Pfeilen, wenn man voraussetzt, dass jede indirekten Pfeil-Verbindung – also jeder Verbindung über eine Zwischenstation – durch eine direkte Verbindung abgekürzt werden kann: wenn man von n über m zu k kommt (d.h. $n \leq m \leq k$), dann kommt man auch direkt von n zu k (d.h. $n \leq k$). Die Relation R_{\leq} ist *transitiv*, eine Eigenschaft, die sie mit $R_{<}$ und R_N teilt.

Wenn man also voraussetzt, dass die Relation R_{\leq} reflexiv und transitiv ist, dann versteht sich die (vollständige) Darstellung von R_N als (unvollständige) Darstellung von R_{\leq} . (Auf dieses spezielle Verhältnis zwischen R_N und R_{\leq} kommen wir in Abschnitt 2.4 zurück.) Eine ähnliche Voraussetzung hatten wir bei dem Hasse-Diagramm auf S. 11 gemacht, das man als verkürzte Darstellung der folgenden Relation R_P verstehen kann. Wie man sich leicht überlegt, ist R_P transitiv und reflexiv; auch hier musste man sich die aus der Reflexivität und Transitivität von R_P ergebenden Pfeile dazudenken.

Reflexivität und Transitivität sind *strukturelle Eigenschaften* zweistelliger Relationen, also Eigenschaften, die sich ausschließlich aus der Konstellation der Pfeile ergeben (im Gegensatz z.B. zu der Eigenschaft, nur zwischen Zahlen zu bestehen). Zweistellige Relationen lassen sich nach ihren strukturellen Eigenschaften klassifizieren. Bevor wir das für unsere Beispiele tun, müssen wir noch weitere solche Eigenschaften kennenlernen.

R_P hat eine Eigenschaft, die sie mit R_N und R_{α} teilt, nicht aber mit $R_{<}$ oder R_{\leq} . Ein Blick auf den oberen Teil des Diagramms zeigt dies:



\emptyset steht in der Teilmengenbeziehung zu allen drei Einermengen, aber natürlich ist keine Teilmenge der anderen. Es gilt also insbesondere weder $\{5\} \subseteq \{8\}$ noch $\{8\} \subseteq \{5\}$. Diese Situation kann bei $R_{<}$ oder R_{\leq} nicht vorkommen, die im Gegensatz zu R_P *konnex* sind: je zwei Elemente ihres Bereichs sind immer durch einen Pfeil verbunden.

Auch R_V ist reflexiv, transitiv und nicht konnex. Dennoch würde eine Pfeil-Diagramm ganz anders aussehen. (Auf die graphische Darstellung von Relationen wie R_V kommen wir in Abschnitt 2.3 zu sprechen.) Das liegt daran, dass R_V im Gegensatz zu den anderen hier betrachteten Relationen *symmetrisch* ist: wenn A so heißt wie B , dann heißt B so wie A . Die anderen Relationen

sind allesamt *antisymmetrisch*: man kann sie nur in dem Fall umkehren, in dem sie zwischen einem Objekt und sich selbst bestehen.

Hier sind die allgemeinen Definitionen für die genannten strukturellen Eigenschaften zweistelliger Relationen:

Definitionen

Sei A eine beliebige Menge und R eine zweistellige Relation über A .

R ist *reflexiv* gdw. für jedes x in R s Bereich gilt: xRx .

R ist *transitiv* gdw. für alle x, y und z gilt: wenn xRy und yRz , so xRz .

R ist *konnex* gdw. für alle x und y in R s Bereich gilt: xRy oder yRx (oder beides).

R ist *symmetrisch* gdw. für alle x und y gilt: wenn xRy , dann yRx .

R ist *antisymmetrisch* gdw. für alle x und y gilt: wenn xRy und yRx , dann ist $x = y$.

AUFGABE

a) Zeigen Sie, dass R_P reflexiv und transitiv ist.

b) Finden Sie eine Relation, die weder symmetrisch noch antisymmetrisch ist.

2.3 Äquivalenzrelationen und Partitionen

Die Relation R_N ist reflexiv, symmetrisch und transitiv. Für diese spezielle Kombination von Eigenschaften gibt es einen Namen:

Definition

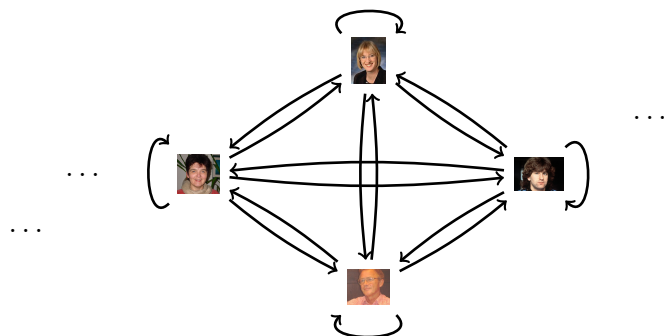
Eine Relation R über einer Menge A ist eine *Äquivalenzrelation* [über A] gdw. R reflexiv, transitiv und symmetrisch ist [und A der Bereich von R ist].

Weitere Beispiele für Äquivalenzrelationen sind:

$$Id_A = \{(x, y) \mid x \in A, y \in A, \text{ und } x = y\} \quad \text{Identität über einer Menge } A$$

$$R_B = \{(A, B) \mid A \text{ hat dieselben Bücher gelesen wie } B\}.$$

Äquivalenzrelationen lassen sich auf verschiedene Weisen darstellen. Eine davon werden wir in diesem Abschnitt kennenlernen, eine weitere im Abschnitt 3. Zunächst einmal schauen wir uns an, wie das Pfeil-Diagramm von R_V aussieht. Der Bereich der Relation ist sehr groß; wir beschränken uns auf einen winzigen Ausschnitt:



Alle vier auf diesem Diagramm eingezeichneten Personen heißen *Eike*. Da jede so heißt wie die andere, führt von jeder ein Pfeil zu jeder anderen. Da jede so heißt wie sie selbst, führt von jeder ein Pfeil zu sich selbst. Wenn wir einmal annehmen, dass niemand anders *Eike* heißt, dann führt von keiner dieser vier Personen ein Pfeil zu irgendeiner anderen Person außerhalb dieser Zeichnung; und keine der vier Personen wird von einem Pfeil von außerhalb erreicht. Die Vier bilden, wie man sagt, eine *Zelle*. Geht man einmal davon aus, dass alle Personen einen Vornamen haben, dann zerlegt offenbar R_V das gesamte Pfeil-Diagramm in lauter einzelne, verschieden große Zellen: eine, in der alle Fritze miteinander verbunden sind, eine mit allen Marias etc.

Intuitiv gesprochen bestehen Äquivalenzrelationen R zwischen Objekten, die etwas gemeinsam haben: den Vornamen, die Identität, die Leseerfahrung etc., also zwischen Objekten die (in einem für R spezifischen Sinn) hinreichend ähnlich (eben *äquivalent*) sind. Dass das so ist, sieht man, wenn man sich Beispiele überlegt. Warum das so ist, werden wir in Abschnitt 3. sehen.

AUFGABE

Sei

$$R_G = \{(A, B) \mid A \text{ und } B \text{ sind Personen mit demselben Geburtstag}\}.$$

Zeigen Sie, dass R_G eine Äquivalenzrelation ist. In wie viele Zellen zerlegt sie die Menge der Personen?

Jede Äquivalenzrelation unterteilt ihren Bereich in Zellen, aus denen keine Pfeile herausführen, in die keine Pfeile hineingehen und innerhalb derer alle Objekte miteinander durch Pfeile verbunden sind. Bei R_B bestehen die Zellen aus Personen, die jeweils dieselben Bücher gelesen haben, bei Id_A besteht jede Zelle aus einem Element von A usw. Man nennt diese Zellen die *Äquivalenzklassen* von R und schreibt

$$|x|_R$$

für die Äquivalenzklasse, in der ein gegebenes Element x zu finden ist, also $\{y \mid xRy\}$. Man kann zeigen, dass nicht nur jede Äquivalenzrelation ihren Bereich in Zellen zerlegt ('partitioniert', wie man auch sagt), sondern dass auch umgekehrt zu jeder beliebigen Zerlegung einer Menge eine entsprechende Äquivalenzrelation passt. Als Vorbereitung dafür dient dieser kleine:

Hilfssatz

Für beliebige Äquivalenzrelationen R über einer Menge A gilt: xRy gdw. $|x|_R = |y|_R$.

Beweis:

Wenn (" \Rightarrow ") xRy , dann gilt wegen der Symmetrie von R auch yRx . Wenn also $z \in |x|_R$, dann haben wir: $yRxRz$ und somit: yRz wegen der Transitivität von R . Also ist $z \in |y|_R$. Damit gilt $|x|_R \subseteq |y|_R$ und (mit vertauschten Rollen von x und y) auch das Gegenteil. – Sei (" \Leftarrow ") $|x|_R = |y|_R$. Zunächst gilt wegen der Reflexivität von R : $y \in |y|_R$, d.h. also auch: $y \in |x|_R$ und deshalb: xRy .

Um nun den angesprochenen Zusammenhang aufzuzeigen, müssen wir ihn überhaupt erst einmal genau formulieren. Dazu dient die folgende:

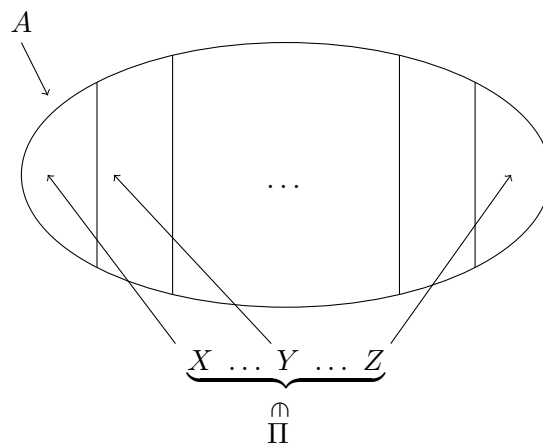
Definition

Eine *Partition* einer Menge A ist eine A 'abdeckende' Menge Π von nicht-leeren, paarweise disjunkten Teilmengen von A , d.h.:

$$\begin{aligned} \Pi &\subseteq \wp(A) \setminus \{\emptyset\}, \\ \bigcup \Pi &= A, \text{ und:} \\ X \cap Y &= \emptyset, \text{ sobald } X, Y \in \Pi \text{ und } X \neq Y \end{aligned}$$

Die Menge $\bigcup \Pi$ ist dabei die *große Vereinigung* über Π , also die Menge aller Objekte, die in einem Element von Π auftauchen:

$$\bigcup \Pi = \{x \mid \text{es gibt ein } X \in \Pi, \text{ so dass gilt: } x \in X\}$$



Wenn also (wie bei R_N) Π aus lauter Zellen besteht, in denen gleichnamige Personen sitzen, dann ist die Menge Π aller Personen, die überhaupt in irgendeiner dieser Zellen sitzen, also (nach unserer Annahme) die Menge aller Personen überhaupt.

Man bezeichnet die Menge der Äquivalenzklassen einer Äquivalenzrelation R über einer Menge A – also: $\{X \in \wp(A) \mid X = |x|_R, \text{ für ein } x \in A\}$ – als die *von R induzierte Partition* Π_R .

Satz 4a

Π_R heißt nicht nur so, Π_R ist auch eine Partition (wenn R eine Äquivalenzrelation über einer Menge A ist).

Beweis:

Man muß zeigen, dass (i) Π_R nicht die leere Menge (als Element) enthält, (ii) dass jedes $x \in A$ in einer Abteilung von Π liegt und dass (iii) zwei verschiedene Elemente von Π_R stets disjunkt sind. (i) folgt aus der Reflexivität von R : xRx gilt immer (für $x \in A$) und impliziert $x \in |x|_R \neq \emptyset$. (ii) gilt ebenfalls wegen der Reflexivität von R : da xRx für alle $x \in A$ gilt, ist auch $x \in |x|_R \in \Pi_R$ für alle $x \in A$. Für (iii) argumentieren wir indirekt und nutzen dabei den obigen Hilfssatz aus: wäre $|x|_R \cap |y|_R \neq \emptyset$ und $|x|_R \neq |y|_R$, müsste irgendein z sowohl in $|x|_R$ als auch in $|y|_R$ liegen. Wir hätten dann also: xRz und yRz . Mit Symmetrie und Transitivität bedeutet dies aber: xRy . Mit dem Hilfssatz folgt dann: $|x|_R = |y|_R$, was der Voraussetzung widerspricht.

Jetzt wissen wir also, dass Äquivalenzrelationen tatsächlich ihre Bereiche zerlegen. Und die Umkehrung gilt auch:

Satz 4b

Jede Partition wird von einer Äquivalenzrelation induziert.

Beweis

Es sei Π eine Partition einer Menge A . Die Idee ist, dass *in der gleichen Zelle sitzen* eine Äquivalenzrelation ist. Wir setzen also:

$$R = \{(x, y) \mid \{x, y\} \subseteq X, \text{ für irgendein } X \in \Pi\}$$

und zeigen, dass R (a) eine Äquivalenzrelation ist und (b) Π induziert.

ad (a): *Reflexivität*: wenn $x \in A$, dann gibt es ein $X \in \Pi$ mit $x \in X$ (weil Π A abdecken muss) und somit gilt: $\{x\} = \{x, x\} \subseteq X \in \Pi$; d.h.: xRx . – *Symmetrie*: wenn xRy , dann ist $\{x, y\} = \{y, x\} \subseteq X \in \Pi$ (für ein X), d.h.: yRx . – *Transitivität*: wenn $xRyRz$, dann gibt es X_1 und X_2 , so dass gilt: $\{x, y\} \subseteq X_1 \in \Pi$ und $\{y, z\} \subseteq X_2 \in \Pi$. Aber dann ist $y \in X_1 \cap X_2$, d.h. $X_1 \cap X_2 \neq \emptyset$ und somit: $X_1 = X_2$, nach dem obigen Hilfssatz. Aber dann ist $\{x, z\} \subseteq X_1 = X_2 \in \Pi$, d.h.: xRz .

ad (b): Wir müssen zeigen, dass $\Pi = \Pi_R$. Sei also (“ \Rightarrow ”) $X \in \Pi$. Dann gibt es ein $x \in X$ (denn $X \neq \emptyset$). Aber $X = |x|_R$: für beliebige $y \in A$ ist $y \in X$ gdw. $\{x, y\} \subseteq X$ (denn $x \in X$) gdw. xRy . Also ist $X \in \Pi_R$. – Sei umgekehrt (“ \Leftarrow ”) $|x|_R$ eine R -Äquivalenzklasse, d.h.: $|x|_R = \{y \mid xRy\} = \{y \mid \{x, y\} \in X, \text{ für ein } X \in \Pi\}$. Da Π eine Partition ist, gibt es genau ein $X \in \Pi$, so dass $x \in X$, sagen wir: X_0 . Also ist $|x|_R = \{y \mid \{x, y\} \subseteq X_0\} = X_0$.

Es gibt also einen (wie man in der Mathematik sagt) ‘ein-eindeutigen’ Zusammenhang zwischen Äquivalenzrelationen und Partitionen: jeder Äquivalenzrelation entspricht genau eine Partiti-

on und umgekehrt. Um eine Äquivalenzrelation graphisch darzustellen, stellt man daher für gewöhnlich die entsprechende Partition dar; denn das geht im allgemeinen sehr viel leichter und übersichtlicher als eine Darstellung per Pfeildiagramm. Statt der vielen Pfeile auf S. 20 unten tun es auch zwei einfache Linien zur Abgrenzung der Zelle der Eikes:



Wegen der engen Verwandtschaft zwischen Äquivalenzrelationen und Partitionen ist es üblich, erstere mit letzteren zu ‘identifizieren’, was bedeutet, dass man in der Praxis nicht immer ausdrücklich zwischen ihnen unterscheidet. So kann man z.B. sagen, dass zwei Objekte äquivalent im Sinne einer gegebenen Partition sind, dass sie in derselben Zelle der Äquivalenzrelation sitzen etc. Diese Art der ‘Identifikation’ strenggenommen verschiedener mengentheoretischer Objekte werden wir noch öfter kennenlernen. Insbesondere werden wir in Abschnitt 3. eine weitere Art von Objekt kennenlernen, die wir mit Äquivalenzrelationen (und Partitionen) identifizieren können. Auf diese Weise wird dann auch klar, warum eine Äquivalenzrelation immer nur zwischen Objekten zu bestehen scheinen, die etwas gemeinsam haben.

AUFGABE

Eine zweistellige Relation R über einer Menge A ist euklidisch, falls für x , y und z gilt: Wenn xRy und xRz , dann ist auch yRz . Zeigen Sie:

- (i) Jede Äquivalenzrelation ist euklidisch.
- (ii) Jede reflexive euklidische Relation ist eine Äquivalenzrelation.
- (iii) Nicht jede euklidische Relation ist transitiv.

2.4 Abschlüsse von Relationen

Wir hatten gesehen, dass man bei einem Pfeildiagramm gelegentlich Pfeile weglassen kann, wenn sie sich aus allgemeinen Eigenschaften der Relation ergeben.

Wenn man weiß, dass die dargestellte Relation reflexiv ist, verstehen sich die Selbstverweise von selbst; wenn man weiß, dass sie transitiv ist, braucht man keine indirekten Verbindungen. Ebenso überlegt man sich leicht, dass man zur Darstellung einer symmetrischen Relation die Pfeile durch Verbindungslinien (oder z.B. Doppelpfeile ‘ \leftrightarrow ’) ersetzen kann. Aber es ist wichtig, dass man beim Weglassen von Pfeilen darauf hinweist, welche von ihnen man sich dazudenken muss; sonst könnte es bei graphischen Darstellungen wie denen weiter oben auf S. 18 zu Verwechslungen zwischen $R_{<}$ und R_{\leq} bzw. R_N und $R_{<}$ kommen. Aber wenn eine entsprechende *Verabredung* über das Weglassen von Pfeilen getroffen wird, dann *gelten* z.B. die Pfeildiagramme von R_N und $R_{<}$ als eindeutige Darstellungen von $R_{<}$ bzw. R_{\leq} . Denn die Pfeile lassen sich auf eindeutige Weise ergänzen.

Der eindeutigen Ergänzung der Diagramme entspricht eine eindeutige *Erweiterung* der dargestellten Relation. Die Nachfolger-Relation R_N kann man auf genau eine Weise zu einer transitiven Relation erweitern, nämlich die Kleiner-Relation $R_{<}$, die sich wiederum auf eindeutige Weise zu einer reflexiven Relation R_{\leq} erweitern lässt. Genauer gesagt, gibt es jeweils genau eine Möglichkeit, die Relation so zu erweitern, dass das Ergebnis einerseits die erwünschte Eigenschaft (Transitivität oder Reflexivität, ...) hat und andererseits keine überflüssigen Pfeile eingefügt wurden (aber natürlich auch keine weggenommen wurden, aber das soll gerade durch den Begriff der Erweiterung zum Ausdruck gebracht werden). Die jeweiligen Erweiterungen sind also in einem naheliegenden Sinne *minimal*. Minimale Erweiterungen einer Relation in diesem Sinne bezeichnet man der Mengenlehre als *Abschlüsse* der betreffenden Relation *bezüglich* der erwünschten Eigenschaft: $R_{<}$ ist der *transitive Abschluss* von R_N , R_{\leq} ist der *reflexive Abschluss* von $R_{<}$ usw. In diesem Abschnitt werden wir diese Begriffe zunächst präzisieren und dann zeigen, wie man die einzelnen Abschlüsse aus einer beliebig vorgegebenen Relation konstruieren kann.

Der hier benutzte Begriff der Erweiterung ist sehr einfach: die Relation $R_{<}$ erweitert R_N in dem Sinne, dass beim Übergang von R_N zu $R_{<}$ alle R_N -Pfeile erhalten bleiben. Im allgemeinen ist eine (zweistellige) Relation R^* eine *Erweiterung* einer (zweistelligen) Relation R , falls gilt: $R \subseteq R^*$. Und dass R^* eine transitive, symmetrische etc. Erweiterung von R ist, heißt natürlich nur, dass R^* transitiv, symmetrisch etc. und zugleich eine Erweiterung von R ist. Aber dieser Begriff reicht nicht aus, um R^* im allgemeinen eindeutig zu charakterisieren: das kartesische Produkt \mathbb{N}^2 über der Menge aller natürlichen Zahlen ist z.B. (wie man sich leicht überlegt) eine transitive Erweiterung von R_N .

Wir brauchen also noch einen Begriff von *Minimalität*. Wir hatten gesagt, dass z.B. R_{\leq} in dem Sinne eine minimale reflexive Erweiterung von $R_{<}$ ist, dass beim Übergang von $R_{<}$ zu R_{\leq} keine überflüssigen Pfeile hinzugefügt werden, also nur solche, die auch wirklich für die Reflexivität des Resultats benötigt werden. Hätten wir z.B. außer den Selbstverweisen noch einen Pfeil von 3 nach 2 hinzugefügt, wäre das Resultat zwar eine reflexive Relation gewesen, aber eben keine

minimale; und durch Hinzunahme von noch mehr Pfeilen (wie etwa in \mathbb{N}^2) hätten wir uns noch weiter von einer minimalen Erweiterung entfernt. Ähnliches lässt sich vom Übergang von R_N zu $R_<$ sagen. Auch hier kommen nur die Pfeile hinzu, die man wirklich braucht, um eine transitive Erweiterung von R_N zu erhalten.

AUFGABE

- a) Charakterisieren Sie die Menge R_{neu} der beim Übergang von R_N zu $R_<$ neu hinzukommenden Pfeile als Relation zwischen Zahlen:

$$R_{neu} = \{(n, m) \mid ???\}$$

- b) Zeigen Sie, dass jeder neue Pfeil, also jedes Paar in R_{neu} wirklich für die Transitivität der Relation benötigt wird. [Tipp: Gehen Sie indirekt vor.]
- c) Geben Sie eine (vom kartesischen Produkt \mathbb{N}^2 verschiedene) transitive Erweiterung von R_{neu} an, die nicht minimal ist.

Das Fazit aus diesen Beobachtungen ist, dass eine minimale Erweiterung mit einer bestimmten Eigenschaft eine solche ist, die nicht mehr Pfeile enthält als alle anderen Erweiterungen mit der entsprechenden Eigenschaft. Nicht mehr Pfeile zu enthalten als eine Relation S heißt für eine Relation R^* , dass $R^* \subseteq S$. Wir gelangen damit zu der folgenden

Definition

R und R^* seien zweistellige Relationen über einer Menge M , und \mathcal{E} sei eine Eigenschaft zweistelliger Relationen.

- R^* ist eine *minimale Relation* (über M) mit der Eigenschaft \mathcal{E} , falls R^* selbst die Eigenschaft \mathcal{E} besitzt und für alle Relationen S (über M) gilt: wenn S die Eigenschaft \mathcal{E} hat, dann ist S eine Erweiterung von R^* (d.h. $R^* \subseteq S$).
- R^* ist der \mathcal{E} -Abschluss von R , wenn R^* eine minimale Erweiterung von R mit der Eigenschaft \mathcal{E} ist.

Um die Definition besser zu verstehen, kann man den Fall $\mathcal{E} = \text{Transitivität}$ betrachten. Der transitive Abschluss einer Relation R ist nach der zweiten Definition eine minimale Erweiterung von R , die transitiv ist, also eine minimale transitive Erweiterung von R . Nach dem ersten Teil der Definition heißt das, dass der transitive Abschluss von R selbst eine transitive Erweiterung von R ist und zugleich Teilmenge jeder transitiven Erweiterung von R . R^* ist also transitiv, eine Obermenge von R , und jeder Pfeil von R^* ist auch in jeder anderen transitiven Erweiterung von R^* zu finden.

Man beachte, dass die Definition nur festlegt, unter welchen Umständen eine gegebene Relation R^* eine bestimmte andere Relation R abschließt. Damit garantiert sie natürlich weder, dass es ein solches R^* überhaupt gibt noch, dass dieses eindeutig bestimmt ist. Letzteres ist allerdings

kein echtes Problem: wenn nämlich sowohl R_0^* als auch R_1^* \mathcal{E} -Abschlüsse von R sind, sind beide insbesondere Erweiterungen von R mit der Eigenschaft \mathcal{E} . Also gilt nach der obigen Definition insbesondere: $R_1^* \subseteq R_0^*$, und $R_0^* \subseteq R_1^*$, d.h. $R_1^* = R_0^*$. Ob es nun allerdings überhaupt einen \mathcal{E} -Abschluss einer Relation gibt, hängt im allgemeinen von \mathcal{E} (und der Relation selbst) ab.

AUFGABE

Wann besitzt eine Relation einen antisymmetrischen Abschluss?

Für manche Eigenschaften \mathcal{E} gibt es allerdings immer einen entsprechenden Abschluss:

Satz 5a - c

Jede zweistellige Relation R über einer Menge M besitzt (genau) einen

$\left. \begin{array}{l} \text{a) reflexiven} \\ \text{b) symmetrischen} \\ \text{c) transitiven} \end{array} \right\}$ Abschluss.

Der Beweis von Satz 5c macht ausgiebig vom sog. *Prinzip der vollständigen Induktion* Gebrauch, nach dem eine Aussage für jede natürliche Zahl gilt, wenn sie für die 0 gilt und sich von jeder Zahl n auf ihren Nachfolger $n + 1$ überträgt:

Für alle Eigenschaften \mathcal{E} gilt:

Wenn: $\mathcal{E}(0)$ "Induktionsanfang (IA)"

und: $\mathcal{E}(n + 1)$, sobald $\mathcal{E}(n)$ "Induktionsschritt (IS)"

Dann: $\mathcal{E}(n)$, für alle natürlichen Zahlen n

(Der unterstrichene Teil des IS heißt "Induktionsvoraussetzung".)

Beweisskizze

ad b): Es sei R eine zweistellige Relation über einer Menge M . Wir konstruieren zunächst die sog. inverse Relation R^{-1} , indem wir die Pfeile in R einfach umdrehen:

$$R^{-1} = \{(x, y) \mid yRx\}.$$

Dann zeigen wir, dass $R \cup R^{-1}$ der symmetrische Abschluss von R ist: jedes Element von R ist auch ein Element von $R \cup R^{-1}$, also ist $R \cup R^{-1}$ eine Erweiterung von R . $R \cup R^{-1}$ ist auch symmetrisch: denn wenn $(x, y) \in R \cup R^{-1}$, dann ist entweder $(x, y) \in R$, und deshalb $(y, x) \in R^{-1} \subseteq R \cup R^{-1}$ oder umgekehrt $(x, y) \in R^{-1}$, und somit $(y, x) \in R \subseteq R \cup R^{-1}$. Und wenn S eine symmetrische Erweiterung von R ist, dann gilt für jedes Element $(x, y) \in R \cup R^{-1}$: entweder gilt xRy und damit auch xSy , weil S eine Erweiterung von R ist; oder es gilt $xR^{-1}y$, und somit yRx – aber dann ist auch ySx (aus demselben Grund) und somit: xSy , denn S ist symmetrisch.

ad c): Im Falle des transitiven Abschlusses (auch als *transitive Hülle* bekannt) konstruiert man zunächst Relationen $R_1, R_2, \dots, R_n, \dots$, die jeweils alle 'R-Ketten' der Länge $\leq n$ enthalten, d.h. Paare (x_1, x_2) , $(x_2, x_3), \dots, (x_{n-1}, x_n)$, so dass $x_1 R x_2 R \dots R x_{n-1} R x_n$. R_1 enthält also die direkten Pfeilverbindungen, R_2 die indirekten mit einer Zwischenstation etc. Die genaue Konstruktion macht Gebrauch vom Prinzip der vollständigen Induktion: wir definieren R_1 und sagen dann für eine beliebige Zahl n , wie man R_{n+1} aus R_n gewinnt:

$$R_1 = R;$$

$$R_{n+1} = R_n \cup \{(x, z) \mid \text{es gibt ein } y, \text{ so dass gilt: } xR_n y R_n z\}.$$

Man überlege sich, dass die Relationen R_n mit wachsendem n immer größer werden: wenn $n \leq m$, dann ist $R_n \subseteq R_m$. (Den Beweis, der ebenfalls vom Induktionsprinzip Gebrauch macht, lassen wir weg.) Jetzt sammeln wir alle diese Verlängerungen von R auf:

$$R^* = \{(x, y) \mid \text{es gibt ein } n, \text{ so dass gilt: } xR_n y\}.$$

R^* ist offensichtlich eine Erweiterung von R . R^* ist auch transitiv: wenn nämlich xR^*yR^*z , dann muss es irgendwelche n und m geben, so dass gilt: $xR_n yR_m z$. Aber eines der beiden, k , ist die größere Zahl oder $i = m = k$. Also gilt für dieses k (s.o.): $xR_k yR_k z$ – und damit: $xR_{k+1} z$. Also ist auch xR^*z , denn $R_{k+1} \subseteq R^*$. Damit bleibt zu zeigen, dass R^* minimal ist. Wenn aber S eine transitive Erweiterung von R ist, dann gilt für jedes n , dass $R_n \subseteq S$: $R_0 \subseteq S$ gilt, weil S eine Erweiterung von $R = R_0$ ist; und wenn $R_n \subseteq S$, dann folgt $R_{n+1} \subseteq S$ aus der Transitivität von S . (WIE?) Da aber aus xR^*y folgt, dass $xR_n y$ für ein k gilt und $R_n \subseteq S$ für jedes n – also auch für k – richtig ist, gilt: $R^* \subseteq S$. [PS: Nach den vorangehenden Erläuterungen hätte man vielleicht eher erwartet, dass in R_{n+1} die R_n -Ketten nur um jeweils einen R -Pfeil verlängert werden:

$$R_{n+1} = R_n \cup \{(x, z) \mid \text{es gibt ein } y, \text{ so dass gilt: } xR_n yRz\}$$

Tatsächlich hätte es diese Konstruktion auch getan, aber der Beweis wäre etwas umständlicher ausgefallen, weil dann man (wohl) nicht ohne einen Hilfssatz ausgekommen wäre, nach dem aus $xR_n yR_m z$ stets folgt, dass $xR_{n+m} z$.]

AUFGABE

Beweisen Sie Satz 5a.

3 Funktionen

3.1 Grundlegendes

Der in diesem Abschnitt einzuführende Funktionsbegriff ist ein zentraler Begriff der Mengenlehre und der modernen Mathematik überhaupt. Daher zunächst ein paar motivierende Worte. Eine Funktion ist, intuitiv gesprochen, ein Objekt – z.B. in der Mathematik in der Regel eine Zahl – in Abhängigkeit von irgendetwas – z.B. einer anderen Zahl oder irgendeinem anderen Objekt. Das mathematische Standardbeispiel ist die aus der Schulmathematik bekannte *Quadratfunktion*; ein Beispiel aus dem Alltagsleben ist der Begriff der *Einwohnerzahl*. In beiden Fällen handelt es sich um eine von einem anderen Objekt abhängige Zahl: was x^2 ist, hängt davon ab, welche Zahl x ist, die Einwohnerzahl hängt vom Ort ab. Die Objekte, von denen die Zahl abhängt, bezeichnet man als die *Argumente* der betreffenden Funktion; die Zahl, die für ein einzelnes Argument herauskommt, ist der *Wert* der Funktion (für dieses Argument). Die Zahl 25 ist demnach der Wert der Quadratfunktion für die Zahl 5 (aber auch für -5), die Argumente der Einwohnerzahl-Funktion sind die Orte, etc. pp. Im allgemeinen sind die Funktionswerte, wie gesagt, nicht immer Zahlen. Zum Beispiel bezeichnet der Begriff *Nationalitätszeichen* eine Funktion, deren Argumente Länder und deren Werte Buchstabenfolgen ('CH' für die Schweiz usw.) sind.

Funktionen stellt man sich am besten als Tabellen vor, in deren Spalten die Funktionswerte den Argumenten zugeordnet werden. Die drei genannten Funktionen sehen danach so aus:

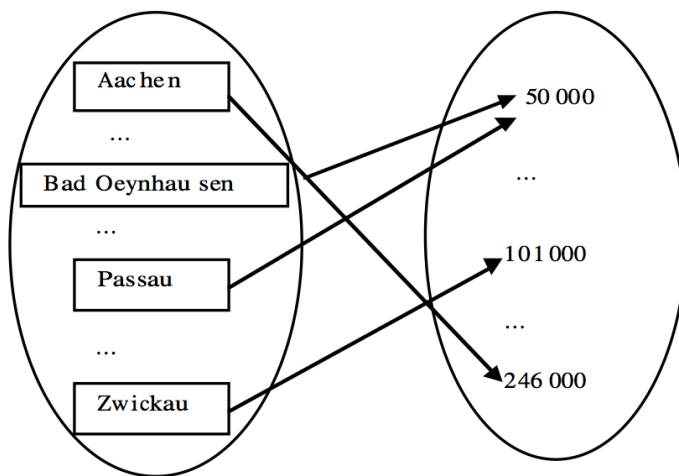
...	...	Aachen	240 000	Afghanistan	AFG
-3	9	Aalen	67 000	Albanien	AL
-2	4	Ahlen	52 000	Algerien	DZ
-1	1
0	0	Bad Oeyenhausen	50 000
1	1
2	4	Passau	50 000
3	9
4	16	Zweibrücken	34 000	Zaire	ZRE
...	...	Zwickau	92 000	Zentralafrikanische Rep.	RCA

Quadratfunktion *Einwohnerzahlfunktion* *Nationalitätszeichenfunktion*

Für die gleich zu liefernde mengentheoretische Präzisierung des allgemeinen Funktionsbegriffs macht man sich am besten einige strukturelle Eigenschaften anhand dieser Beispiele klar. Zunächst einmal ist zu beobachten, dass die Reihenfolge, in der die Argumente in der Funktionstabelle (auch als *Wertetafel* bekannt) offenbar keine Rolle spielt. Wir haben hier konventionelle Reihenfolgen (nach Größe oder Alphabet) gewählt, aber für die Quadratfunktion an sich macht es z.B. keinen Unterschied, in welcher Reihenfolge die Zahlen genannt werden. (Strenggenommen kann man die Tabelle sowieso nicht vollständig hinschreiben, weil es viel zu viele Zahlen gibt, womit sich die Frage der Reihenfolge erübrigt!) Wichtig ist nun allerdings, dass jedes Argument

nur einmal vorkommt. Denn hätte Göttingen z.B. zwei Einträge, wäre die Tabelle entweder redundant (weil dieselbe Einwohnerzahl zweimal erscheint) oder widersprüchlich (weil Göttingen dann zwei Einwohnerzahlen haben müsste).

Neben der Tabellenform gibt es auch die Darstellung einer Funktion durch Pfeile. Üblicherweise gibt man in dem Fall die Menge der Argumente – den *Definitionsbereich* der Funktion – getrennt von der Menge der Werte – den *Wertebereich* – an, selbst wenn es sich (wie im Falle der Quadratfunktion) um dieselbe Menge handelt. Eine Pfeildarstellung der Einwohnerzahlfunktion sieht z.B. so aus:



Diese Pfeil-Darstellung verdeutlicht die Hauptcharakteristika von Funktionen. Danach wird jedem Element des Definitionsbereichs ein (und nur ein) Element des Wertebereichs zugeordnet, wobei es zwar vorkommen kann, dass ein Wert – in diesem Falle die Zahl 50 000 – mehr als einmal zugeordnet wird, niemals dagegen ein Argument mehr als einen Wert zugewiesen bekommt.

Die Pfeildarstellung legt nahe, den Funktionsbegriff in ähnlicher Weise wie den der Relation zu präzisieren:

Definition

Eine *Funktion* von einer Menge A in eine Menge B ist eine Teilmenge f des kartesischen Produkts $A \times B$, so dass es für jedes $x \in A$ genau ein $y \in B$ gibt, für das gilt: $(x, y) \in f$.

NB: Eine Funktion von A nach B ist immer auch eine zweistellige Relation über $A \cup B$. Alle für Relationen definierten Begriffe und Notationen übertragen sich damit auf Funktionen.

Notationskonventionen

Statt 'f ist eine Funktion von A nach B' schreibt man kürzer: ' $f : A \rightarrow B$ '. Wenn $f : A \rightarrow B$ und $x \in A$, dann bezeichnet ' $f(x)$ ' dasjenige y , so dass $(x, y) \in f$.

Wenn also f die Nationalitätszeichenfunktion, S die Menge aller Staaten und F die Menge aller Buchstabenfolgen ist, dann gilt: $f : S \rightarrow F$, und $f(\text{Liechtenstein}) = \text{'FL'}$ (schon gewusst?).

AUFGABE

Zeigen Sie, dass $f : A \rightarrow B$ und $f : A' \rightarrow B'$ impliziert, dass $A = A'$.

Zeigen Sie, dass $f : A \rightarrow B$ und $f : A' \rightarrow B'$ nicht impliziert, dass $B = B'$.

Zeigen Sie: Wenn $f : A \rightarrow B$, $g : A \rightarrow B'$, und $f(x) = g(x)$ für alle $x \in A$, dann ist $f = g$.

Wenn $f : A \rightarrow B$, ist also A eindeutig bestimmt. B ist dagegen nicht eindeutig bestimmt; aber es gibt immer eine minimale Wertemenge:

Definition

Wenn $f : A \rightarrow B$ heißt A der *Definitionsbereich* von f . Notation: $\text{dom}(f)$.

Wenn $f : A \rightarrow B$ heißt $\{y \in B \mid \text{es gibt ein } x, \text{ so dass } f(x) = y\}$ der *Wertebereich* von f .

Notation: $\text{rge}(f)$; $f : A \xrightarrow{\text{auf}} B$.

Es gilt also für beliebige Funktionen $f: f : \text{dom}(f) \rightarrow \text{rge}(f)$.

Die Abkürzung ' $f : A \xrightarrow{\text{auf}} B$ ' liest man als: ' f ist eine Funktion von A auf B ', was also heißt, dass jedes Element von B mindestens einmal als Wert fungiert. Man sagt in diesem Falle auch, dass die Funktion f *surjektiv* auf B ist. Jede Funktion ist demnach surjektiv auf ihren Wertebereich.

3.2 Eigenschaften von Funktionen

Am Beispiel der (gerundeten) Einwohnerzahlen hatten wir gesehen, dass bei einer Funktion die Pfeile zusammenlaufen können. Sie müssen es aber nicht, wie das Beispiel der Länderkennzeichen zeigt. Im letzteren Falle spricht man von einer *injektiven* Funktion:

Definition

Eine Funktion f ist *injektiv*, falls es keine voneinander verschiedene Elemente x und x' des Definitionsbereichs von f gibt, so dass gilt: $f(x) = f(x')$.

Notation: $f : A \xrightarrow{1-1} B'$

Oft findet man die Injektivitätsbedingung positiv ausgedrückt: $f(x) = f(x')$ impliziert $x = x'$.

Es ist oft zweckmäßig, injektive Funktionen als Funktionen auf ihren Wertebereich zu betrachten. Die folgende Begriffsbildung unterstreicht dies:

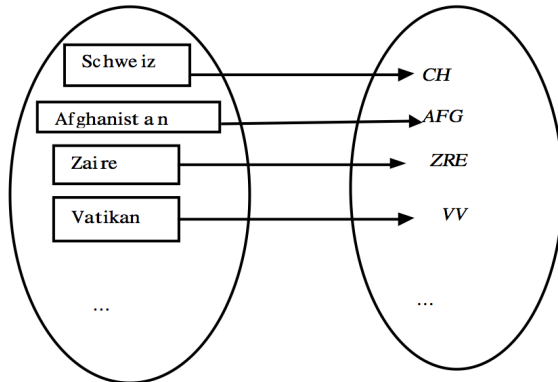
Definition

Eine *Bijektion* zwischen einer Menge A und einer Menge B ist eine injektive Funktion mit Wertebereich B . Notation: $f : A \xrightarrow{\text{auf}} B$.

Man beachte, dass zwar nicht jede Funktion eine Bijektion ist, wohl aber jede injektive Funktion.

Das einfachste Beispiel für eine Bijektion ist die schon als zweistellige Relation betrachtete Identität über einer Menge $A : Id_A = \{(x, x) \mid x \in A\}$. Ein weiteres Beispiel zeigt, wie man sich

Bijektionen im allgemeinen vorzustellen hat. Da niemals zwei Länder dasselbe Nummernschild haben, ist die Nationalitätskennzeichenfunktion v injektiv. Der Wertebereich von v ist die Menge aller Buchstabenfolgen, die überhaupt als Kennzeichen fungieren. Insbesondere landet also bei jedem dieser Buchstabenfolgen ein v -Pfeil. Wegen der Injektivität von v geht auch bei jedem Land nur ein Pfeil los. Länder und Buchstabenfolgen (im Wertebereich von v) lassen sich also so untereinander anordnen, dass jeweils ein Pfeil jeweils ein Land mit einer Abkürzung verbindet:



Zwei Dinge fallen auf. Zum einen kann so eine Zuordnung nur funktionieren, wenn Definitionsbereich und Wertebereich *gleich viele Elemente* besitzen. In der Mengenlehre wird diese Tatsache ausgenutzt, um den Begriff der Anzahl auf unendlich große Mengen zu verallgemeinern; wir kommen darauf zurück. Eine andere Auffälligkeit von Bijektionen ist, dass man sie ‘umdrehen’ kann und dann offenbar wieder eine Bijektion erhält:

Satz 6

Es sei $f : A \xrightarrow[\text{auf}]{1-1} B$. Dann ist $f^{-1} : B \xrightarrow[\text{auf}]{1-1} A$.

Zur Erinnerung: f^{-1} war die zu f inverse Relation, also $\{(y, x) \mid f(x) = y\}$.

Beweis

Drei Dinge sind zu zeigen: (i) dass f^{-1} eine Funktion von B nach A ist; (ii) dass f^{-1} injektiv ist; (iii) dass f^{-1} surjektiv auf A ist.

ad (i): Da f surjektiv auf B ist, gibt es für jedes $y \in B$ ein $x \in A$, so dass $(x, y) \in f$, d.h. $(y, x) \in f^{-1}$. Es bleibt zu zeigen, dass es nie mehr als ein solches x gibt. Wenn aber $(y, x) \in f^{-1}$ und $(y, x') \in f^{-1}$, dann ist $f(x) = f(x')$ und somit $x = x'$, denn f ist injektiv.

ad (ii): $f^{-1}(y) = f^{-1}(y')$ bedeutet, dass es ein $x \in A$ gibt, so dass $(y, x) \in f^{-1}$ und $(y', x) \in f^{-1}$, d.h.: $(x, y) \in f$ und $(x, y') \in f$. Aber dann ist $y = y'$, denn f ist eine Funktion.

ad (iii): Da $f : A \xrightarrow[\text{auf}]{1-1} B$, gibt es insbesondere für jedes $x \in A$ ein $y \in B$, so dass $(x, y) \in f$, d.h. es gibt für jedes $x \in A$ ein $y \in B$, so dass $(y, x) \in f^{-1}$.

Da jede injektive Funktion eine Bijektion auf ihren Wertebereich ist, erhalten wir das folgende:

Korollar Es sei $f : A \xrightarrow{1-1} B$. Dann ist $f^{-1} : \text{rge}(f) \xrightarrow[\text{auf}]{1-1} A$.

Neben der Inversenbildung ist die sog. *Verknüpfung* von Funktionen eine häufig benötigte Operation. Dabei handelt es sich um das ‘Hintereinanderausführen’ zweier Funktionen auf denselben Ausgangswert. So kann man zum Beispiel eine Zahl zuerst quadrieren und dann das Ergebnis durch 2 teilen; oder man bestimmt zunächst das Kennzeichen eines Landes und dann die Anzahl dessen Buchstaben. In jedem Fall kann man sich diese Hintereinanderausführung zweier Operationen (= Funktionen) als eine einzige, komplexe Operation vorstellen: *die Hälfte des Quadrats* bzw. *die Länge des Nationalitätskennzeichens*. Die folgende Begriffsbildung erfasst genau diese Idee:

Definition

A , B und C seien irgendwelche Mengen, so dass gilt: $f : A \rightarrow B$, und $g : B \rightarrow C$. Dann ist $g \circ f = \{(x, z) \mid z = g(f(x))\}$.

Es ist klar, dass in diesem Fall $g \circ f : A \rightarrow C$.

Wenn also q die Quadratfunktion ist und h die Funktion, die jeder Zahl die Hälfte dieser Zahl zuordnet, dann ist $h \circ q$ die Hälfte-des-Quadrats-Funktion. Man beachte, dass $h \circ q \neq q \circ h$, denn z.B. ist $h \circ q(2) = h(q(2)) = h(4) = 2 \neq 1 = q(1) = q(h(2)) = q \circ h(2)$. Wenn dagegen v die Nationalitätskennzeichenfunktion ist und λ für jede (endliche) Folge von Buchstaben die Länge angibt, ist $\lambda \circ v$ die Länge des Kennzeichens – also z.B. $\lambda \circ v(\text{Schweiz}) = 2$ – während es $v \circ \lambda$ gar nicht gibt, denn v lässt sich nur auf Länder anwenden, während es sich bei den Werten von λ um Zahlen handelt. Man sieht, dass man zwei beliebige Funktionen nicht unbedingt miteinander verknüpfen kann. Der folgenden Satz betrifft nur solche Funktionen, deren Verknüpfung auch tatsächlich existiert:

Satz 7

f , g und h seien Funktionen. Dann gilt:

- | | | |
|-----|---|----------------------------------|
| (a) | $(f^{-1})^{-1} = f$ | |
| (b) | $f \circ Id_{\text{dom}(f)} = Id_{\text{range}(f)} \circ f = f$ | |
| (c) | $f \circ f^{-1} = f^{-1} \circ f = Id_{\text{range}(f)}$ | ...wenn f injektiv ist |
| (d) | $(f \circ g) \circ h = f \circ (g \circ h)$ | ...wenn es diese Funktionen gibt |

Beweis

<u>ad (a)</u>	$(f^{-1})^{-1} = \{(x, y) \mid (y, x) \in f^{-1}\}$	Definition von ' \circ^{-1} '
	$= \{(x, y) \mid (x, y) \in f\}$	Definition von ' \circ^{-1} '
	$= f$	Enxtensionalität

<u>ad (b)</u>	Sei $x \in \text{dom}(f), z \in \text{rge}(f)$. Dann gilt:	
	$(x, z) \in f \circ \text{Id}_{\text{dom}(f)}$ gdw. $z = f(\text{Id}_{\text{dom}(f)}(x))$	Definition von ' \circ '
	gdw. $z = f(x)$	Definition von $\text{Id}_{\text{dom}(f)}$
	gdw. $f(x) = z$	sowieso
	gdw. $\text{Id}_{\text{rge}(f)}(f(x)) = z$	Definition von $\text{Id}_{\text{rge}(f)}$
	gdw. $(x, z) \in \text{Id}_{\text{rge}(f)} \circ f$	Definition des Funktionswerts

Also sind $f \circ \text{Id}_{\text{dom}(f)}$ und $\text{Id}_{\text{rge}(f)} \circ f$ dieselbe Funktion

<u>ad (c)</u>	$(x, y) \in f \circ f^{-1}$ gdw. $y = f^{-1}(f(x))$	Definition von ' \circ '
	gdw. $(f(x), y) \in f^{-1}$	Definition des Funktionswerts
	gdw. $(y, f(x)) \in f$	Definition von ' \circ^{-1} '
	gdw. $f(y) = f(x)$	Definition des Funktionswertes
	gdw. $y = x$	Injektivität von f

Somit ist $f \circ f^{-1}$ die Identität, sobald f injektiv ist. Außerdem ist nach (a):

$f^{-1} \circ f = f \circ (f^{-1})^{-1}$, also auch die Identität, denn f^{-1} ist injektiv, wenn f injektiv ist.

<u>ad (d)</u>	$(f \circ g) \circ h(x) = (f \circ g)(h(x))$	Definition von ' \circ '
	$= f(g(h(x)))$	Definition von ' \circ '
	$= f(g \circ h(x))$	Definition von ' \circ '
	$= f \circ (g \circ h)(x)$	Definition von ' \circ '

Die in Satz 7 genannten Gleichungen erinnern an einige der Rechenregeln für die Mengenoperationen (welche?) und bilden einen der Ausgangspunkte der modernen Algebra (Gruppentheorie).

Wir beschließen unseren Durchgang durch die allgemeinsten Eigenschaften von Funktionen mit dem schon angekündigten Zusammenhang zu Äquivalenzrelationen und Partitionen. Zunächst machen wir dazu die folgende:

Beobachtung

Denselben Funktionswert zu besitzen ist eine Äquivalenzrelation.

Genauer: Wenn $f : A \rightarrow B$, ist $\{(x, y) \in A^2 \mid f(x) = f(y)\}$ eine Äquivalenzrelation über A . R nennt man die *Wertgleichheit* unter f .

Das liegt an den entsprechenden Eigenschaften der Identität (als zweistelliger Relation): $f(x) = f(x)$ (Reflexivität); $f(y) = f(x)$, sobald $f(x) = f(y)$ (Symmetrie); und $f(x) = f(z)$, sobald $f(x) = f(y) = f(z)$.

Die Beobachtung macht deutlich, dass das Bestehen einer Äquivalenzrelation zwischen zwei Objekten oft auf einer Gemeinsamkeit zwischen ihnen zu beruhen scheint. Im Falle der Wertgleichheit ist das Gemeinsame der Funktionswert. Das Interessante ist nun, dass sich *jede* Äquivalenzrelation als Wertgleichheit unter einer Funktion verstehen lässt. Wenn also R reflexiv,

symmetrisch und transitiv ist, besteht R zwischen zwei Objekten x und y , wenn sie denselben f -Wert einer bestimmten Funktion f besitzen. In den meisten konkreten Beispielen lässt sich so ein f einigermaßen leicht angeben. Die Relation der Gleichaltrigkeit ist z.B. die Wertgleichheit der Altersfunktion, also der Funktion, die jeder Person ihr Alter (z.B. in Jahren) zuweist; die weiter oben betrachtete Relation R_B , die gleichen Bücher gelesen zu haben, lässt sich als Wertgleichheit der Funktion f_B auffassen, die jeder Person die Menge der von ihr gelesenen Bücher zuweist; usw. Jetzt zeigen wir, dass diese Beispiele keine Einzelfälle sind; dabei machen wir ganz wesentlich von dem in Kap. 2. dargestellten Zusammenhang zwischen Äquivalenzrelationen und Partitionen Gebrauch:

Satz 8

Wenn R eine Äquivalenzrelation über einer Menge A ist, gibt es eine Menge B und eine Funktion f_R von A nach B , so dass R die Wertgleichheit zu f_R ist.

Beweis

Der ist erstaunlich einfach: man nimmt als B einfach die von R induzierte Partition Π_R von A und setzt $f_R := \{(a, |a|_R) \mid a \in A\}$. Dass R eine Funktion von A nach B ist, ist klar. Gezeigt werden muss, dass R die Wertgleichheit zu f_R ist, dass also für beliebige Elemente a und b von A gilt: aRb gdw. $f_R(a) = f_R(b)$, was nach der Definition von f_R heißt, dass $|a|_R = |b|_R$. Aber das hatten wir schon bewiesen – als Vorbereitung zum Beweis von Satz 4!

Man beachte, dass die im Satz genannte Funktion f_R nicht eindeutig bestimmt ist. Im Falle der Altersgleichheit etwa könnte man ja statt der Äquivalenzklassen Gleichaltriger genausogut natürliche Zahlen, z.B. das Alter der jeweiligen Personen, nehmen; welche Zahlen man dafür nimmt, ist egal, solange gleichaltrige Personen dieselbe und verschieden alte Personen verschiedene Zahlen zugewiesen bekommen.

In Fällen, in denen man nicht so genau weiß, welche Objekte man den Personen als Gemeinsamkeit zuordnen soll, hilft die im Beweis benutzte Methode weiter. Man betrachte z.B. die Relation G , die zwischen Personen a und b besteht, wenn a und b denselben Geschmack haben. Eine entsprechende Funktion f_G , für die G gerade die Wertgleichheit ist, müsste (intuitiv gesprochen) jeder Person ihren Geschmack zuordnen. Aber was ist der Geschmack einer Person? Es handelt sich offenbar um irgendetwas Abstraktes, schwer Fassbares, von dem wir höchstens wissen, wann sich zwei Personen ihn teilen. Wenn Letzteres der Fall ist und es sich – wie jedenfalls die Redeweise vom gleichen Geschmack suggeriert – bei G tatsächlich um eine Äquivalenzrelation handelt, können wir die Konstruktion des obigen Beweises ausnutzen, in dem wir den Geschmack einer Person a einfach mit der Äquivalenzklasse $|a|_G$ identifizieren. Das garantiert einerseits, dass die Existenz eines ansonsten recht diffus und dubios erscheinenden Objekts (as Geschmack) zumindest im Rahmen der Mengenlehre außer Frage steht, ohne dass andererseits das, was wir über den Geschmack von Personen wissen, mit dieser Identifikation unvereinbar wäre. Wollten wir z.B. die Geschmäcker verschiedener Personen vergleichen oder bewerten, liefe dies – die genannte Identifikation vorausgesetzt – auf Vergleiche und Bewertungen von Äquivalenzklassen von in ihren Geschmacksurteilen ununterscheidbaren Personen hinaus. Diese Art der Konstruktion abstrakter Gegenstände geht auf den Begründer der modernen Logik, Frege, zurück und wird als *Abstraktion durch Äquivalenzklassenbildung*

bezeichnet. Sie spielt z.B. in der Wissenschaftstheorie, aber auch vor allem in der Mathematik, eine große Rolle. So kann man z.B. die Ladung von Partikeln als Menge von ladungsgleichen Partikeln auffassen, die Richtung einer Geraden als die Menge der zu ihr parallelen Geraden (das war eines von Freges Originalbeispielen) oder – wie wir noch genauer sehen werden – die Größe einer Menge als die Klasse der ihr gleichgroßen Mengen.

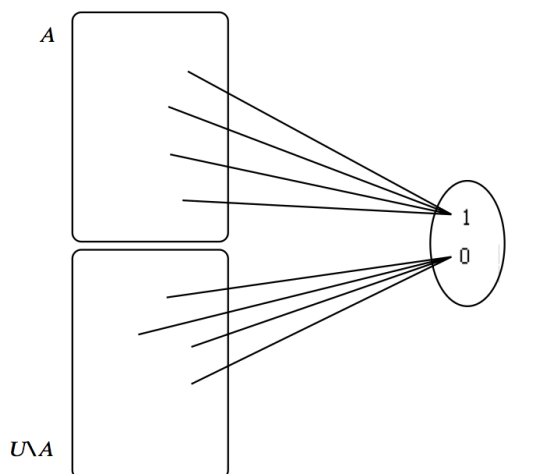
3.3 Charakteristische Funktionen

Wie schon im Zusammenhang mit Mengenoperationen werden wir jetzt eine beliebige Menge U festhalten, die wir als das *Universum* (bestehend aus *Individuen*) ansehen wollen. Den Teilmengen A von U , also den Elementen von $\wp(U)$, entsprechen nun gewisse Funktionen, die die Elemente von U danach sortieren, ob sie in A sind oder nicht: wenn ja, weisen sie die Zahl 1 zu, wenn nein, bekommt das Element von U die 0 zugeordnet:

Definition

Sei $A \subseteq U$. Eine Funktion $f : U \rightarrow \{0, 1\}$ charakterisiert A (relativ zu U), falls für alle $x \in U$ gilt: $f(x) = 1$ gdw. $x \in A$.

Es gilt: $f = \{(x, 1) \mid x \in A\} \cup \{(x, 0) \mid x \in U \setminus A\}$.



Bemerkung:

Falls $A \subseteq U$, gibt es genau ein f das A relativ zu U charakterisiert; umgekehrt charakterisiert jedes $f : U \rightarrow \{0, 1\}$ genau eine Teilmenge von U .

Man bezeichnet solche f mit Definitionsbereich U und Werten in $\{0, 1\}$ als *charakteristische Funktionen* (über U). Charakteristische Funktionen spielen in der Semantik eine große Rolle, wo sie sich oft als zweckmäßiger erweisen als die charakterisierten Mengen, zu denen sie in einer Eins-zu-eins-Beziehung stehen. Doch damit nicht genug. Auch *Relationen* über einem gegebenen Universum U werden durch entsprechende Funktionen ersetzt. Um zu sehen, wie das geht,

betrachten wir am besten ein Beispiel. Sei U die Menge der Personen und B die (zweistellige) Relation der Bruderschaft, d.h.: $B = \{(x, y) \mid \text{die Person } x \text{ ist Bruder der Person } y\}$. Für jede einzelne Person y sei nun B_y die Menge von y s Brüdern. xB_y gilt genau dann, wenn $x \in B_y$. Da die B_y Teilmengen von U sind, lassen sie sich durch Funktionen f_y charakterisieren. Es gilt also: xB_y genau dann, wenn $f_y(x) = 1$. Da es für jedes y so ein f_y gibt, bildet die Menge der Paare (y, f_y) eine Funktion g von U in die Menge der charakteristischen Funktionen über U , und es gilt: $g(y)(x) = 1$ gdw. $f_y(x) = 1$ gdw. xB_y , d.h. g enthält alle Informationen über B . Dieses g wird in der Semantik als Ersatz für die Relation B benutzt. Den Übergang von B zu g bezeichnet man als *Schönfinklung* oder *Currying* – nach den Begründern der sog. kombinatorischen Logik, in der diese “Kodierungstechnik” eine zentrale Rolle spielt: Moses Schönfinkel und Haskell B. Curry. Hier ist die allgemeine:

Definition

Sei R eine zweistellige Relation über einer Menge U , und $f : U \rightarrow [U \rightarrow \{0, 1\}]$. f schönfinkelt R (relativ zu U), falls für alle $y \in U$ und $x \in U$ gilt:

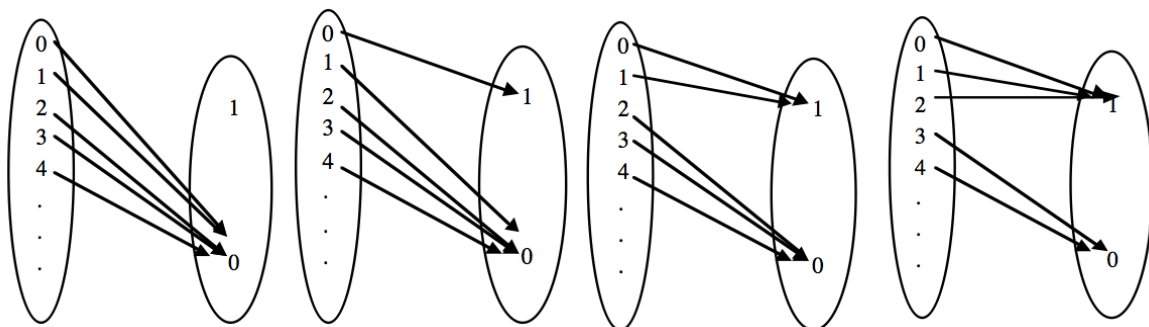
$$f(y)(x) = 1 \text{ gdw. } xRy.$$

(Die Notation $f : U \rightarrow [U \rightarrow \{0, 1\}]$ drückt dabei aus, dass die Werte von f selbst wieder Funktionen von U nach $\{0, 1\}$ sind.)

Ein konkretes Beispiel hilft, sich diese abstrakten Begriffe klarzumachen. Es sei R die Kleiner-Beziehung zwischen den natürlichen Zahlen, d.h.:

$$R = \{(n, m) \mid n \text{ und } m \text{ sind natürliche Zahlen, und } n < m\}.$$

Wie sieht die Schönfinklung von R aus? Zunächst bestimmt man dazu für jedes m die Menge M_m der ‘ R -Vorgänger’, d.h. $\{n \mid n < m\}$. Z.B. ist $M_0 = \emptyset$, $M_1 = \{0\}$, $M_2 = \{0, 1\}$, etc. pp. (Man beachte, dass M_m stets m Elemente enthält!) Jede dieser Mengen M_m lässt sich nun durch eine Funktion f_m charakterisieren, die gerade den Elementen von M_m die 1 und allen anderen Zahlen die 0 zuordnet. Zum Beispiel sehen $f_0 - f_3$ so aus:



Jedes f_m sagt also für eine gegebene Zahl n , ob n kleiner ist als m . Die Schönfinklung der

Kleiner-Beziehung R ist nun einfach die Funktion g , die jeder Zahl m ihre entsprechende Funktion f_m zuordnet: $g(0)$ ist die Funktion ganz links, dann kommt $g(1)$ usw.

AUFGABEN

- a) Zeigen Sie, dass jede Funktion $f : U \rightarrow [U \rightarrow \{0, 1\}]$ genau eine zweistellige Relation über U schönfinkelt und dass jede zweistellige Relation über U von genau einer Funktion schöngefinkelt wird.
- b) Zeigen Sie, dass die Schönfinklung der Kleiner-Beziehung injektiv ist. Geben Sie ein Beispiel für eine zweistellige Relation, deren Schönfinklung nicht injektiv ist.

Die Schönfinkerei lässt sich auch auf mehrstellige Relationen verallgemeinern. Wir verzichten hier auf die (umständliche) Darstellung dieses Sachverhalt, werden aber in *Semantik I* nicht darum herumkommen.

3.4 Indizierungen und Folgen

Funktionen f können dazu benutzt werden, um sich mit den Elementen einer Menge I auf die Elemente einer anderen Menge M zu beziehen. Die $i \in I$ dienen dann quasi als Labels oder Namen für die $m \in M$. In diesem Falle nennt man f auch eine *Indizierung* von M (mit einer *Indexmenge* I). Ein typisches Beispiel sind numerische Indizierungen, wie man sie in vielen alltäglichen Bereichen verwendet. Ein handelsüblicher Fernseher verfügt z.B. über eine (virtuelle und programmierbare) Sendertabelle, in der Zahlen Fernsehsender zugeordnet sind: $1 \rightarrow \text{ARD}$, $2 \rightarrow \text{ZDF}$, etc. Bei einer solchen Zuordnung handelt es sich um eine Indizierung s , also eine (übrigens oft nicht einmal injektive) Funktion, deren Zweck es ist, die einzelnen Funktionswerte (Sender) über entsprechende Argumente (Zahlen) zu identifizieren (anzuklicken). In so einem Fall schreibt man die Funktionswerte gern – wie der Name *Indizierung* schon andeutet – mit Hilfe von Indizes: s_1, s_2, s_3 , usw. statt $s(1), s(2), s(3), \dots$. Und wenn man s selbst bezeichnen will, nennt man die Indexmenge oft mit und schreibt: $(s_i)_{i \in I}$, wobei I die Menge der beim Fernseher einstellbaren Zahlen (z.B. $\{0, \dots, 99\}$) ist.

Wie in diesem Beispiel eignen sich natürliche Zahlen – und vor allem *Anfangsstücke* I von \mathbb{N} , also Mengen der Form $\{0, 1, 2, \dots, n\}$ – wegen ihrer (eben: natürlichen) Ordnung dazu, beliebige Gegenstandsbereiche zu indizieren und dabei gleichzeitig anzuordnen. Da für die Indexmenge eine Ordnungsstruktur vorausgesetzt werden darf, überträgt sich letztere auf die indizierte Menge. Man schreibt dann die Indizierung auch als: $\langle m_i \rangle_{i \leq n}$ oder $\langle m_0, m_1, \dots, m_n \rangle$ und nennt sie eine $n + 1$ -stellige (*endliche*) *Folge*. $n + 1$ -stellige Folgen sind insofern wie $n + 1$ -Tupel, als auch für sie gilt: $\langle a_0, a_1, \dots, a_n \rangle = \langle b_0, b_1, \dots, b_n \rangle$ gdw. für alle i zwischen 0 und n gilt: $a_i = b_i$, aber sie können nicht alle Aufgaben der $n + 1$ -Tupel übernehmen: als Funktionen sind sie Mengen von geordneten Paaren und setzen also schon den Paarbegriff voraus! Im Unterschied zu Tupeln können Folgen auch unendlich lang sein, d.h. von der Gestalt $\langle a_0, a_1, \dots, a_n, a_{n+1}, \dots \rangle$, was man auch als $\langle a_i \rangle_{i \rightarrow \omega}$ oder $\langle a_i \rangle_{i \in \mathbb{N}}$ schreibt. Die Folge bricht dann also niemals ab, und sie bestimmt sich einzig und allein durch die Reihenfolge ihrer Glieder. Übrigens: wie alle Funktionen sind Indizierungen und Folgen nicht notwendigerweise injektiv, d.h. einzelne Glieder können mehr als einmal auftreten.

3.5 Unendlichkeit und Kardinalität

Wir hatten beobachtet, dass man offenbar nur dann eine Bijektion zwischen zwei Mengen finden kann, wenn diese gleich viele Elemente haben. Dieser Umstand wird in der Mengenlehre dazu genutzt, den Begriff der Anzahl von Elementen einer Menge – ja, den Begriff der Zahl überhaupt – zu definieren (oder, wie man sagt: ‘mit mengentheoretischen Mitteln zu *rekonstruieren*’). Wir definieren dazu zunächst eine Beziehung zwischen beliebigen Mengen:

Definition: Es seien A und B Mengen. Dann gilt:

A und B sind genau dann gleichmächtig – symbolisch: $A \sim B$ – wenn es eine Bijektion von A nach B gibt.

Satz 9

Für beliebige Mengen A , B und C gilt:

- a) $A \sim A$;
- b) wenn $A \sim B$, dann auch $B \sim A$;
- c) wenn $A \sim B$ und $B \sim C$, dann auch $A \sim C$.

Beweis:

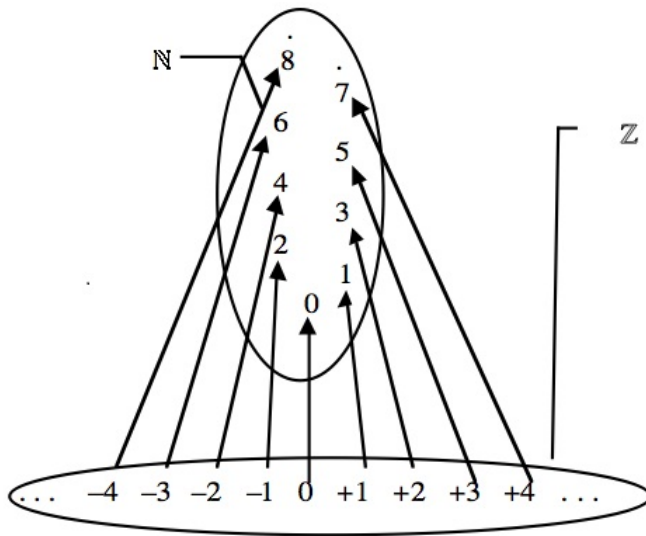
a) gilt, weil Id_A eine Bijektion ist.

b) gilt, weil nach Satz 6 f^{-1} eine Bijektion von B nach A ist, wenn f eine Bijektion von A nach B ist.

c) gilt, weil $f' \circ f$ eine Bijektion von A nach C ist, wenn f und f' Bijektionen von A nach B bzw. von B nach C sind; den Beweis schenken wir uns.

\sim ist demnach eine Äquivalenzrelation auf dem Universum aller Mengen. Wie sehen die Äquivalenzklassen aus? Im endlichen Bereich handelt es sich jeweils um alle Mengen, die eine bestimmte Anzahl von Elementen haben: eine Äquivalenzklasse enthält nur die leere Menge (als einzige Menge mit 0 Elementen), eine andere alle Einermengen, etc. Frege hat diese Äquivalenzklassen als logische Rekonstrukte der natürlichen Zahlen betrachtet. Die Konstruktion ist außer Mode geraten, weil sie sich im Rahmen der axiomatischen Mengenlehre in dieser Form nicht nachvollziehen lässt.

Im endlichen Bereich sind rasch Beispiele für Gleichmächtigkeit gefunden: $\{0, 1, 2\} \sim \{3, 5, 6\}$, $\{a, b\} \sim \{a\}$, wenn $a = b$ usw. – was bestätigt, dass Gleichmächtigkeit im endlichen Falle auf gleiche Elementanzahl hinausläuft. Aber die Definition funktioniert auch im Unendlichen. Z. B. ist $\mathbb{N} \sim \mathbb{Z}^+ [= \mathbb{N} \setminus \{0\}]$, denn man kann die beiden Mengen durch ‘Verschieben’ aufeinander abbilden, indem man jeder natürlichen Zahl ihren Nachfolger zuordnet; denn die in Abschnitt 2.2 betrachtete Relation $R_{\mathbb{N}}$ ist in der Tat eine Bijektion zwischen \mathbb{N} und \mathbb{Z}^+ . Man beachte, dass diese Bijektion eine Menge (\mathbb{N}) mit einer ihrer *echten* Teilmengen (\mathbb{Z}^+) in eine Eins-zu-eins-Beziehung setzt, was nur im Unendlichen funktioniert. (Das ist eine der sogenannten *Paradoxien* des Unendlichen). Doch es kommt noch toller. Denn die echte Teilmenge kann indem Sinne unendlich viel kleiner sein, als das Komplement selber wieder unendlich ist. Ein Beispiel ist: $\mathbb{Z} \sim \mathbb{N}$. Eine entsprechende Bijektion sieht so aus:



Sogar die rationalen Zahlen lassen sich bijektiv auf die positiven ganzen Zahlen abbilden. Wir führen die Idee (die vom Begründer der Mengenlehre, Georg Cantor, stammt) hier nur anhand der positiven Brüche vor; der Rest wird sich aus allgemeinen Überlegungen ergeben. Um eine Bijektion zu finden, ordnet man die (positiven) Brüche in einer unendlichen Tabelle an:

		Nenner							
		1	2	3	4	5	6	...	
{	Zähler	0	0/1	0/2	0/3	0/4	0/5	0/6	...
	1	1/1	1/2	1/3	1/4	1/5	
	2	2/1	2/2	2/3	2/4	
	3	3/1	3/2	3/3	
	4	4/1	4/2	
	5	5/1	
...		

Aber \mathbb{Z}^+ lässt sich auch in einer solchen Tabelle anordnen:

		Nenner						
		1	2	3	4	5	6	...
Zähler =	0	1	2	4	7	11	16	...
	1	3	5	8	12	17
	2	6	9	13	18
	3	10	14	19
	4	15	20
	5	21

Wir fangen dabei links oben mit dem Zählen an und arbeiten uns langsam nach rechts unten durch, Diagonale für Diagonale. Offenkundig stehen die beiden Tabellen in einer Eins-zu-eins-Beziehung und somit auch die positiven Brüche und die positiven ganzen Zahlen.

Angesichts des letzten Beispiels liegt der Verdacht nahe, dass man zwischen zwei unendlichen Mengen immer eine Bijektion finden kann, dass sich also Größengleichheit zwischen unendlichen Mengen – wenn überhaupt – jedenfalls nicht mit Gleichmächtigkeit definieren lässt, weil man so keine Abstufungen im Unendlichen bekäme. Dieser Eindruck ist grottenfalsch:

Satz von Cantor

Keine Menge ist gleichmächtig mit ihrer Potenzmenge.

Beweis

Angenommen, es gäbe eine Menge A , so dass $A \sim \wp(A)$ und somit auch eine Bijektion f von A nach $\wp(A)$. Wir setzen $X = \{a \in A \mid a \notin f(a)\}$. Da $X \in \wp(A)$ und f surjektiv ist, muss es ein $x \in A$ geben, so dass $f(x) = X$. Aber dann wäre $x \in f(x)$ gdw. $x \in \{a \in A \mid a \notin f(a)\}$ gdw. $x \notin f(x)$, was nicht sein kann.

Für den Fall $A = \mathbb{N}$, lässt sich der Beweis auch mit Hilfe charakteristischer Funktionen veranschaulichen. Wir hatten gesehen, dass die Teilmengen einer Menge - also die Elemente der Potenzmenge - in einer Eins-zu-eins-Beziehung zu den charakteristischen Funktionen über dieser Menge stehen. Die Teilmengen der natürlichen Zahlen werden durch Funktionen von \mathbb{N} nach $\{0, 1\}$ charakterisiert und entsprechen demnach unendlichen Folgen (im Sinne von Abschnitt 3.5) von Nullen und Einsen: die Menge \mathbb{N} selbst entspricht der Folge $\langle 1, 1, 1, \dots \rangle$ aus lauter Einsen, weil jede Zahl von der charakteristischen Funktion den Wert 1 zugewiesen bekommt; die leere Menge entspricht der Folge $\langle 0, 0, 0, \dots \rangle$ aus lauter Nullen, weil die charakteristische Funktion von \emptyset stets die 0 zuordnet; die Einermenge $\{1\}$ der Folge $\langle 0, 1, 0, 0, 0, \dots \rangle$, die nur an der zweiten Stelle die Zahl 1 zuordnet, denn die 1 steht an der zweiten Stelle der natürlichen Zahlen (nach der 0) und nur der 1 ordnet diese Funktion die 1 zu; die Menge der geraden Zahlen entspricht der Folge $\langle 0, 0, 1, 0, 1, 0, 1, 0, 1, \dots \rangle$, die - außer ganz am Anfang (weil 0 nicht gerade ist) immer zwischen 0 und 1 alterniert; etc. pp. Angenommen, wir könnten die Potenzmenge von \mathbb{N} aufzählen, also in eine Eins-zu-eins-Beziehung zu \mathbb{N} selbst setzen. Schematisch sähe eine solche Aufzählung so aus:

	0	1	2	3	4	5	...
0	w_0^0	w_1^0	w_2^0	w_3^0	w_4^0	w_5^0	...
1	w_0^1	w_1^1	w_2^1	w_3^1	w_4^1	w_5^1	...
2	w_0^2	w_1^2	w_2^2	w_3^2	w_4^2	w_5^2	...
3	w_0^3	w_1^3	w_2^3	w_3^3	w_4^3	w_5^3	...
4	w_0^4	w_1^4	w_2^4	w_3^4	w_4^4	w_5^4	...
5	w_0^5	w_1^5	w_2^5	w_3^5	w_4^5	w_5^5	...
...

Dabei ist die erste Zeile die Folge, mit der Nummer 0 (also der Wert der hypothetischen Bijektion für das Argument 0), dann kommt Folge Nummer 1 usw. Jede Folge Nummer n besteht selbst wieder aus dem Wert für die 0 (w_0^n), gefolgt von dem für die 1 (w_1^n) etc. – wobei jeder dieser Werte entweder 0 oder 1 ist. Im Cantorschen Beweis passiert nun folgendes: wir betrachten zunächst die Diagonalfolge, also die Folge von Nullen und Einsen, die gerade auf den Punkten der Diagonale erscheinen und *drehen diese Werte um*: wo auf der obigen Diagonalen 0 stand, schreiben wir eine 1, wo eine 1 war, steht jetzt eine 0:

	0	1	2	3	4	5	...
0	$\overline{w_0^0}$	w_1^0	w_2^0	w_3^0	w_4^0	w_5^0	...
1	w_0^1	$\overline{w_1^1}$	w_2^1	w_3^1	w_4^1	w_5^1	...
2	w_0^2	w_1^2	$\overline{w_2^2}$	w_3^2	w_4^2	w_5^2	...
3	w_0^3	w_1^3	w_2^3	$\overline{w_3^3}$	w_4^3	w_5^3	...
4	w_0^4	w_1^4	w_2^4	w_3^4	$\overline{w_4^4}$	w_5^4	...
5	w_0^5	w_1^5	w_2^5	w_3^5	w_4^5	$\overline{w_5^5}$...
...

Die hervorgehobenen Werte $\overline{w_j^i}$ stehen dabei gerade immer gerade für das Gegenteil des ursprünglichen Wertes w_j^i (also für $1 - w_j^i$!). Diese Werte bilden offenbar wieder eine (unendliche) Folge von Nullen und Einsen, also eine charakteristische Funktion einer Teilmenge von \mathbb{N} . Da wir angenommen hatten, dass die Tabelle jede charakteristische Funktion erfasst, muss sie auch diese erfassen. Aber das kann nicht sein: wenn eine Folge auf der Tabelle die Nummer n hat (d.h. in der n ten Zeile steht), stimmt die hervorgehobene ‘Antifolge’ an der n ten Stelle nicht mit ihr überein!

4 Automaten und formale Sprachen

4.1 Ketten und Sprachen

Gegeben sei eine endliche, nicht-leere Menge Σ ('Alphabet'), die keine Funktionen (als Elemente) enthält.

Eine *Kette* über Σ ist eine endliche Folge über Σ , d.h. eine Funktion, deren Definitionsbereich ein (endliches) Anfangstück $\{0, 1, \dots, n\}$ der natürlichen Zahlen ist und deren Werte Elemente von Σ sind. Σ^* ist die Menge aller Ketten über Σ ('Kleenescher Stern'). Die Verkettungsoperation ist diejenige (zweistellige) Funktion $\hat{\cdot} : (\Sigma^* \times \Sigma^*) \rightarrow \Sigma^*$, so dass für alle $x, y \in \Sigma^*$ gilt:

$$x \hat{\cdot} y = x \cup \{(i + |x|, y_i) \mid i \in \text{dom}(y)\},$$

d.h. die resultierende Kette fängt mit x an und fügt dann die Glieder von y an.

Wenn $x \in \Sigma^*$, ist die *Länge von x* – geschrieben: $|x|$ – die Anzahl der Elemente (des Definitionsbereichs) von x .

Notationskonventionen

- Für den (möglichen) Fall $|x| = 0$ ist gibt es nur die leere Kette, die als ' ε ' (statt ' \emptyset ') notiert wird.
- Für die Menge $\Sigma^* \setminus \{\varepsilon\}$ der nicht-leeren Ketten über Σ schreiben wir auch ' Σ^+ '.
- Wir schreiben Ketten (x_0, \dots, x_n) einfach als ' $x_0 \dots x_n$ ', also ohne Klammern, Kommas und Zwischenräume.
- Für den Fall $|x| = 1$ identifizieren wir sogar die Ketten x mit ihrem Wert x_0 und schreiben z.B. auch ' $x \in \Sigma$ ', obwohl strenggenommen nur gilt: $x_0 \in \Sigma$.

Bemerkung

Für alle $x, y, z \in \Sigma^*$ gilt:

- $x \hat{\cdot} \varepsilon = \varepsilon \hat{\cdot} x = x$,
d.h. (algebraisch gesprochen): ε ist (links- und rechts-) *neutral* bzgl. der Verkettungsoperation.
- $(x \hat{\cdot} y) \hat{\cdot} z = x \hat{\cdot} (y \hat{\cdot} z)$,
d.h. (algebraisch gesprochen): die Verkettungsoperation ist *assoziativ*.

Die zweite Bemerkung rechtfertigt die folgende

Notationskonvention

Wir schreiben meist ' xy ' statt ' $x \hat{\cdot} y$ '.

Eine *Sprache* L über Σ ist eine Teilmenge von Σ^* : $L \subseteq \Sigma^*$.

NB: \emptyset und $\{\varepsilon\}$ sind Sprachen über Σ^* – und es gilt: $\emptyset \neq \{\varepsilon\}$!

Für jede natürliche Zahl $n (\geq 0)$ und jedes $x \in \Sigma^*$ lässt sich die n -fache Potenz von x durch folgende Induktion definieren:

IA: $x^0 = \varepsilon$

IS: $x^{n+1} = x^n \hat{\cdot} x$

4.2 Ersetzungssyntaxen

Eine *Ersetzungssyntax* G ist ein Quadrupel $G = (N, \Sigma, R, S)$, so dass gilt:

- N und Σ sind voneinander disjunkte Alphabete: $N \cup \Sigma = \emptyset$; Kategoriennamen N und Wörter Σ
- $R \subseteq (N \cup \Sigma)^* \times (N \cup \Sigma)^*$; Regeln
- $S \in N$. Satzsymbol

Notationskonvention

Wenn $(l, r) \in R$, schreiben wir ' $l \rightarrow r$ ' statt ' (l, r) '.

Wenn $G = (N, \Sigma, R, S)$ eine Ersetzungssyntax ist und $\rho = l \rightarrow r \in R$, dann heißt ein Paar der Gestalt $(u \hat{\ } l \hat{\ } v, u \hat{\ } r \hat{\ } v) \in [(N \cup \Sigma)^*]^2$ eine *Anwendung* der Regel ρ ; $y \in (N \cup \Sigma)^*$ ist unmittelbar G -herleitbar aus $x \in (N \cup \Sigma)^*$, wenn es eine Regel $\rho \in R$ gibt, so dass (x, y) eine Anwendung von ρ ist.

Eine *Herleitung* gemäß einer Ersetzungssyntax G ist eine endliche Folge (x_1, \dots, x_n) von Ketten über Σ , so dass gilt: wenn $1 \leq i \leq n-1$, ist (x_i, x_{i+1}) eine Anwendung einer Regel $r_i \in R$. y ist G -herleitbar aus x heißt, dass es eine Ableitung (gem. G) der Gestalt (x_1, \dots, x_n) gibt, so dass $x_1 = x$ und $x_n = y$.

Bemerkung:

G -Herleitbarkeit ist die transitive Hülle der unmittelbaren G -Herleitbarkeit.

Notation

- ... für unmittelbare G -Herleitbarkeit von y aus x : $x \xrightarrow{1} y$
- für G -Herleitbarkeit von y aus x : $x \xrightarrow{*} y$

Eine G -Ableitung (x_1, \dots, x_n) ist eine Herleitung gem. G , für die gilt: $x_1 = S$; die Ableitung *terminiert*, falls außerdem $x_n \in \Sigma^*$. Die von einer Ersetzungssyntax G *erzeugte Sprache* $L(G)$ ist die Menge aller $x \in \Sigma^*$, für die es eine (terminierende) G -Ableitung (x_1, \dots, x_n) gibt mit $x_n = x$.

4.3 Kontextfreiheit

Eine Ersetzungssyntax $G = (N, \Sigma, R, S)$ ist *kontextfrei*, falls jedes $\rho \in R$ entweder die Gestalt $S \rightarrow \varepsilon$ hat oder von der Form $A \rightarrow r$ ist, wobei $A \in N$ und $r \neq \varepsilon$.

Beispiele

- $G_0 = (N_0, \Sigma_0, R_0, S_0)$ ist kontextfrei, wobei:

- $N_0 = \{S_0\}$
- $\Sigma_0 = \{a, b\}$
- $R_0 = \{S_0 \rightarrow \varepsilon, S_0 \rightarrow S_0 S_0, S_0 \rightarrow ab\}$

Es gilt: $L(G_0) = \{(ab)^n \mid n \geq 0\}$.

- $G_1 = (N_1, \Sigma_1, R_1, S_1)$ ist kontextfrei, wobei:

- $N_1 = \{A, B, S_1\}$
- $\Sigma_1 = \{a, b\}$
- $R_1 = \{S_1 \rightarrow \varepsilon, S_1 \rightarrow AS_1B, A \rightarrow a, B \rightarrow b\}$

Es gilt: $L(G_1) = \{a^n b^n \mid n \geq 0\}$.

- $G_2 = (N_2, \Sigma_2, R_2, S_2)$ [= Fuchs' Syntax] ist kontextfrei, wobei:
 - $N_2 = \{A, B, C, S_2\}$
 - $\Sigma_2 = \{a, b, c\}$
 - $R_2 = \{S_2 \rightarrow AS_2C, S_2 \rightarrow B, S_2 \rightarrow \varepsilon, A \rightarrow a, B \rightarrow b, B \rightarrow Bb, C \rightarrow c\}$
[= ρ_1, \dots, ρ_7]

Es gilt: $L(G_2) = \{a^n b^m c^n \mid n, m \geq 0\}$.

4.4 Normalität

Eine kontextfreie Ersetzungssyntax $G = (N, \Sigma, R, S)$ ist *normal*, falls jedes $\rho \in R$ entweder die Gestalt $S \rightarrow \varepsilon$ hat oder von der Form $A \rightarrow BC$ oder von der Form $D \rightarrow e$ ist, wobei $A, B, C, D \in N$ und $e \in \Sigma$.

Beispiele

- $G_3 = (N_3, \Sigma_3, R_3, S_3)$ ist normal, wobei:
 - $N_3 = \{S_3, A, B\}$
 - $\Sigma_3 = \{a, b\}$
 - $R_3 = \{S_3 \rightarrow \varepsilon, S_3 \rightarrow S_3 S_3, S_3 \rightarrow AB, A \rightarrow a, B \rightarrow b\}$

Es gilt: $L(G_3) = \{(ab)^n \mid n \geq 0\} = L(G_0)$.

- $G_4 = (N_4, \Sigma_4, R_4, S_4)$ ist normal, wobei:
 - $N_4 = \{A, B, X, S_4\}$
 - $\Sigma_4 = \{a, b\}$
 - $R_4 = \{S_4 \rightarrow \varepsilon, S_4 \rightarrow AX, X \rightarrow S_4 B, A \rightarrow a, B \rightarrow b\}$

Es gilt: $L(G_4) = \{a^n b^n \mid n \geq 0\} = L(G_1)$.

Anders als andere Beschränkungen des Regelformats wirkt sich Normalität nicht auf die Beschreibbarkeit von Sprachen (als Mengen von Zeichenketten) aus:

Normalisierungssatz

Für jede kontextfreie Syntax $G = (N, \Sigma, R, S)$ gibt es eine normale Syntax $G^+ = (N^+, \Sigma, R^+, S)$, so dass gilt: $L(G) = L(G^+)$, d.h. G und G^+ sind *schwach äquivalent*.

4.5 Regularität

Eine kontextfreie Ersetzungssyntax $G = (N, \Sigma, R, S)$ ist *rechtsregulär* [bzw. *linksregulär*], falls jedes $\rho \in R$ entweder die Gestalt $S \rightarrow \varepsilon$ hat oder von der Form $A \rightarrow bC$ [bzw. $A \rightarrow Cb$] oder von der Form $D \rightarrow e$ ist, wobei $A, B, C, D \in N$ und $b, c, e \in \Sigma$.

Beispiele

- $G_5 = (N_5, \Sigma_5, R_5, S_5)$ ist rechtsregulär, wobei:
 - $N_5 = \{S_0, C\}$

- $\Sigma_5 = \{a, b\}$
- $R_5 = \{S_5 \rightarrow \varepsilon, S_5 \rightarrow aC, C \rightarrow bS_5\}$

Es gilt: $L(G_5) = \{(ab)^n \mid n \geq 0\} = L(G_0) = L(G_3)$.

- $G_6 = (N_6, \Sigma_6, R_6, S_6)$ ist linksregulär, wobei:

- $N_6 = \{S_0, C\}$
- $\Sigma_6 = \{a, b\}$
- $R_6 = \{S_6 \rightarrow \varepsilon, S_6 \rightarrow Cb, C \rightarrow S_6a\}$

Es gilt: $L(G_6) = \{(ab)^n \mid n \geq 0\} = L(G_0) = L(G_3) = L(G_5)$.

Die Beispiele legen einen Verdacht nahe – der sich bestätigt:

Beobachtung

Jede rechtsreguläre Grammatik ist schwach äquivalent zu einer linksregulären Grammatik und umgekehrt.

Grund: Wir geben wieder nur die Konstruktion an; wenn $G = (N, \Sigma, R, S)$ rechtsregulär ist, lässt sich ein äquivalentes linksreguläres $G^l = (N, \Sigma, R^l, S)$ definieren durch:

$$\begin{aligned} R^l &= \{S \rightarrow Aa \mid A \rightarrow a \in R\} \\ &\cup \{A \rightarrow Ba \mid B \rightarrow aA \in R, B \neq S\} \\ &\cup \{A \rightarrow a \mid S \rightarrow aA \in R\} \\ &\cup \{S \rightarrow x \mid S \rightarrow x \in R, |x| \leq 1\}. \end{aligned}$$

Die letzte Teilmenge betrifft Regeln der Form $S \rightarrow \varepsilon$ oder $S \rightarrow a$ (wobei $a \in \Sigma$), die von R nach R^l übertragen werden. Für linksreguläre $G = (N, \Sigma, R, S)$ wird ein entsprechendes rechtsreguläres $G^r = (N, \Sigma, R^r, S)$ definiert durch:

$$\begin{aligned} R^r &= \{A \rightarrow a \mid S \rightarrow Aa \in R\} \\ &\cup \{B \rightarrow aA \mid A \rightarrow Ba \in R, B \neq S\} \\ &\cup \{S \rightarrow aA \mid A \rightarrow a \in R\} \\ &\cup \{S \rightarrow x \mid S \rightarrow x \in R, |x| \leq 1\}. \end{aligned}$$

Definition

Eine Sprache $L \subseteq \Sigma^*$ über einem Alphabet Σ ist *regulär*, falls es eine (rechts- oder linksreguläre) Ersetzungssyntax G gibt, so dass $L = L(G)$. L ist *kontextfrei*, falls es ein kontextfreies G gibt, so dass $L(G) = L$.

NB: Jede reguläre Sprache ist kontextfrei; die Umkehrung gilt aber nicht. Denn anders als Normalität schränkt Regularität die beschreibbaren Sprachen ein; so ist z.B. $L(G_1) = \{a^n b^n \mid n \geq 0\}$ kontextfrei, aber nicht regulär. Genauso wenig ist jede Sprache kontextfrei; so ist $\{a^n b^n c^n \mid n \geq 0\}$ nachweislich nicht kontextfrei. Die Beweise für diese Sachverhalte würden hier allerdings zu weit führen.

4.6 Nicht-deterministische Automaten

Links- und rechtsreguläre Grammatiken lassen sich durch einfach Konstellationen von Pfeilen zwischen ihren nicht-terminalen Symbolen darstellen, die den Zuständen gewisser Automaten entsprechen:

Ein (*nicht-deterministischer endlicher*) Automat ist ein Quintupel $M = (Q, \Sigma, \Delta, s, F)$, wobei gilt:

- Q ist eine Menge (die Menge der Zustände von M);
- Σ ist ein (terminales) Alphabet;
- $\Delta \subseteq (Q \times \Sigma) \times Q$ (Übergangsrelation, indizierten Pfeilen entsprechend);
- $s \in F$ (Anfangszustand);
- $F \subseteq Q$ (akzeptierende/End- Zustände)

Die Relation Δ lässt sich von einzelnen Wörtern auf Ketten erweitern:

$$\begin{aligned}\Delta_0 &= \{(q, \varepsilon), q\} \mid q \in Q\}; \\ \Delta_{n+1} &= \{(q, xa), q'\} \mid (q, x, q'') \in \Delta_n, (q'', a, q') \in \Delta_n, \text{ für ein } q'' \in Q\}, \text{ wenn } x \in \Sigma^*, a \in \Sigma; \\ \Delta^* &= \cup \Delta_n.\end{aligned}$$

Also ist $\Delta^* \subseteq (Q \times \Sigma^*) \times Q$. Die Doppelklammerung wird im folgenden weggelassen.

Man beachte (einfache Induktion):

- Wenn $n \leq m$, ist $\Delta_n \subseteq \Delta_m$;
- Wenn $(q, x, q') \in \Delta^*$, dann ist $(q, x, q') \in \Delta_{|x|}$.

Wenn $M = (Q, \Sigma, \Delta, s, F)$ ein nicht-deterministischer Automat ist, ist

$$L(M) = \{x \subseteq \Sigma^* \mid (s, x, q) \in \Delta^*, \text{ für ein } q \in F\}$$

die von M beschriebene Sprache. Statt ' $x \in L(M)$ ' sagt man auch: M akzeptiert x .

Ein Automat $M = (Q, \Sigma, \delta, s, F)$ ist *deterministisch*, falls $\delta : (Q \times \Sigma) \rightarrow Q$. (Man beachte die (konventionelle) Kleinschreibung!)

Beobachtung: Wenn $M = (Q, \Sigma, \delta, s, F)$ deterministisch ist, ist $\sigma^* : (Q \times \Sigma^*) \rightarrow Q$.
Beweis per Induktion über die Konstruktion von σ^* (also die Länge der Ketten!)