

Theoretische Informatik I
Logik und Mengenlehre
— Wintersemester 2013/2014 —
DHBW Mannheim

Prof. Dr. Karl Stroetmann

15. Januar 2014

Inhaltsverzeichnis

1	Einleitung	4
1.1	Motivation	4
1.2	Überblick über den Inhalt der Vorlesung	5
2	Mathematische Grundlagen	7
2.1	Prädikatenlogische Formeln	7
2.1.1	Warum Formeln	7
2.1.2	Formeln als Kurzschreibweise	8
2.1.3	Beispiele für Terme und Formeln	11
2.2	Mengen und Relationen	12
2.2.1	Erzeugung von Mengen durch explizites Auflisten	13
2.2.2	Die Menge der natürlichen Zahlen	14
2.2.3	Das Auswahl-Prinzip	14
2.2.4	Vereinigungs-Mengen	15
2.2.5	Schnitt-Menge	15
2.2.6	Differenz-Mengen	15
2.2.7	Potenz-Mengen	16
2.2.8	Bild-Mengen	17
2.2.9	Kartesische Produkte	17
2.2.10	Gleichheit von Mengen	18
2.2.11	Rechenregeln für das Arbeiten mit Mengen	19
2.3	Binäre Relationen	19
2.3.1	Binäre Relationen und Funktionen	19
2.3.2	Binäre Relationen auf einer Menge	26
3	Die Programmier-Sprache SetlX	37
3.1	Einführende Beispiele	37
3.2	Darstellung von Mengen	41
3.3	Paare, Relationen und Funktionen	43
3.4	Allgemeine Tupel	45
3.5	Spezielle Funktionen und Operatoren auf Mengen	45
3.5.1	Anwendung: <i>Sortieren durch Auswahl</i>	48
3.6	Kontroll-Strukturen	48
3.6.1	Schleifen	50

3.6.2	Fixpunkt-Algorithmen	53
3.6.3	Verschiedenes	55
3.7	Fallstudie: Berechnung von Wahrscheinlichkeiten	55
3.8	Fallstudie: Berechnung von Pfaden	57
3.8.1	Berechnung des transitiven Abschlusses einer Relation	58
3.8.2	Berechnung der Pfade	61
3.8.3	Der Bauer mit dem Wolf, der Ziege und dem Kohl	63
3.9	Terme und Matching	66
3.9.1	Konstruktion und Manipulation von Termen	66
3.9.2	Matching	68
3.9.3	Ausblick	71
4	Grenzen der Berechenbarkeit	72
4.1	Das Halte-Problem	72
4.1.1	Informale Betrachtungen zum Halte-Problem	72
4.1.2	Formale Analyse des Halte-Problems	73
4.2	Unlösbarkeit des Äquivalenz-Problems	76
5	Aussagenlogik	78
5.1	Motivation	78
5.2	Anwendungen der Aussagenlogik	79
5.3	Formale Definition der aussagenlogischen Formeln	80
5.3.1	Syntax der aussagenlogischen Formeln	80
5.3.2	Semantik der aussagenlogischen Formeln	81
5.3.3	Extensionale und intensionale Interpretationen der Aussagenlogik	83
5.3.4	Implementierung in SETLX	83
5.3.5	Eine Anwendung	86
5.4	Tautologien	88
5.4.1	Testen der Allgemeingültigkeit in SETLX	89
5.4.2	Nachweis der Allgemeingültigkeit durch Äquivalenz-Umformungen	91
5.4.3	Berechnung der konjunktiven Normalform in SETLX	95
5.5	Der Herleitungs-Begriff	100
5.5.1	Eigenschaften des Herleitungs-Begriffs	102
5.6	Das Verfahren von Davis und Putnam	103
5.6.1	Vereinfachung mit der Schnitt-Regel	104
5.6.2	Vereinfachung durch Subsumption	105
5.6.3	Vereinfachung durch Fallunterscheidung	105
5.6.4	Der Algorithmus	106
5.6.5	Ein Beispiel	106
5.6.6	Implementierung des Algorithmus von Davis und Putnam	107
5.7	Das 8-Damen-Problem	110

6	Prädikatenlogik	117
6.1	Syntax der Prädikatenlogik	117
6.2	Semantik der Prädikatenlogik	121
6.2.1	Implementierung prädikatenlogischer Strukturen in SETLX	125
6.3	Normalformen für prädikatenlogische Formeln	130
6.4	Unifikation	134
6.5	Ein Kalkül für die Prädikatenlogik	138
6.6	<i>Prover9</i> und <i>Mace4</i>	144
6.6.1	Der automatische Beweiser <i>Prover9</i>	144
6.6.2	<i>Mace4</i>	145
7	<i>Prolog</i>	148
7.1	Wie arbeitet <i>Prolog</i> ?	151
7.1.1	Die Tiefensuche	156
7.2	Ein komplexeres Beispiel	158
7.3	Listen	160
7.3.1	Sortieren durch Einfügen	162
7.3.2	Sortieren durch Mischen	164
7.3.3	Symbolisches Differenzieren	165
7.4	Negation in <i>Prolog</i>	171
7.4.1	Berechnung der Differenz zweier Listen	171
7.4.2	Semantik des Negations-Operators in PROLOG	172
7.4.3	Extralogische Prädikate	173
7.5	Die Tiefen-Suche in <i>Prolog</i>	173
7.5.1	Die Bekehrung der Ungläubigen	178
7.6	Der Cut-Operator	181
7.6.1	Verbesserung der Effizienz von <i>Prolog</i> -Programmen durch den Cut-Operator	183
7.7	Literaturhinweise	186

Kapitel 1

Einleitung

In diesem einführenden Kapitel möchte ich zunächst motivieren, warum wir in der Informatik nicht darum herum kommen uns mit den formalen Methoden der mathematischen Logik zu beschäftigen. Anschließend gebe ich einen Überblick über den Rest der Vorlesung.

1.1 Motivation

Informationstechnische Systeme (im folgenden kurz als IT-Systeme bezeichnet) gehören zu den komplexesten Systemen, die die Menschheit je entwickelt hat. Das läßt sich schon an dem Aufwand erkennen, der bei der Erstellung von IT-Systemen anfällt. So sind im Bereich der Telekommunikations-Industrie IT-Projekte, bei denen mehr als 1000 Entwickler über mehrere Jahre zusammenarbeiten, die Regel. Es ist offensichtlich, dass ein Scheitern solcher Projekte mit enormen Kosten verbunden ist. Einige Beispiele mögen dies verdeutlichen.

1. Am 9. Juni 1996 stürzte die Rakete Ariane 5 auf ihrem Jungfernflug ab. Ursache war ein Kette von Software-Fehlern: Ein Sensor im Navigations-System der Ariane 5 misst die horizontale Neigung und speichert diese zunächst als Gleitkomma-Zahl mit einer Genauigkeit von 64 Bit ab. Später wird dieser Wert dann in eine 16 Bit Festkomma-Zahl konvertiert. Bei dieser Konvertierung trat ein Überlauf ein, da die zu konvertierende Zahl zu groß war, um als 16 Bit Festkomma-Zahl dargestellt werden zu können. In der Folge gab das Navigations-System auf dem Datenbus, der dieses System mit der Steuerungs-Einheit verbindet, eine Fehlermeldung aus. Die Daten dieser Fehlermeldung wurden von der Steuerungs-Einheit als Flugdaten interpretiert. Die Steuer-Einheit leitete daraufhin eine Korrektur des Fluges ein, die dazu führte, dass die Rakete auseinander brach und die automatische Selbstzerstörung eingeleitet werden mußte. Die Rakete war mit 4 Satelliten beladen. Der wirtschaftliche Schaden, der durch den Verlust dieser Satelliten entstanden ist, lag bei mehreren 100 Millionen Dollar.

Ein vollständiger Bericht über die Ursache des Absturzes des Ariane 5 findet sich im Internet unter der Adresse

<http://www.ima.umn.edu/~arnold/disasters/ariane5rep.html>

2. Die Therac 25 ist ein medizinisches Bestrahlungs-Gerät, das durch Software kontrolliert wird. Durch Fehler in dieser Software erhielten 1985 mindestens 6 Patienten eine Überdosis an Strahlung. Drei dieser Patienten sind an den Folgen dieser Überdosierung gestorben, die anderen wurden schwer verletzt.

Einen detaillierten Bericht über diese Unfälle finden Sie unter

http://courses.cs.vt.edu/~cs3604/lib/Therac_25/Therac_1.html

3. Im ersten Golfkrieg konnte eine irakische *Scud* Rakete von dem *Patriot* Flugabwehrsystem aufgrund eines Programmier-Fehlers in der Kontrollsoftware des Flugabwehrsystems nicht abgefangen werden. 28 Soldaten verloren dadurch ihr Leben, 100 weitere wurden verletzt.

<http://www.ima.umn.edu/~arnold/disasters/patriot.html>

4. Im Internet finden Sie unter

<http://www.cs.tau.ac.il/~nachumd/horror.html>

eine Auflistung von schweren Unfällen, die auf Software-Fehler zurückgeführt werden konnten.

Diese Beispiele zeigen, dass bei der Konstruktion von IT-Systemen mit großer Sorgfalt und Präzision gearbeitet werden sollte. Die Erstellung von IT-Systemen muß auf einer wissenschaftlich fundierten Basis erfolgen, denn nur dann ist es möglich, die Korrektheit solcher Systeme zu *verifizieren*, also mathematisch zu beweisen. Diese wissenschaftliche Basis für die Entwicklung von IT-Systemen ist die Informatik, und diese hat ihre Wurzeln in der mathematischen Logik. Zur präzisen Definition der Bedeutung logischer Formeln benötigen wir die Mengenlehre. Daher sind die Logik und die Mengenlehre die beiden Gebiete, mit denen wir uns hauptsächlich in dieser einführenden Vorlesung beschäftigen werden. Sowohl die Mengenlehre als auch die Logik haben unmittelbare praktische Anwendungen in der Informatik.

1. Die Schnittstellen komplexer Systeme können mit Hilfe logischer Formeln exakt spezifiziert werden. Spezifikationen, die mit Hilfe natürlicher Sprache erstellt werden, haben gegenüber formalen Spezifikationen den Nachteil, dass sie oft mehrdeutig sind.
2. Die korrekte Funktionsweise digitaler Schaltungen kann mit Hilfe automatischer Beweiser verifiziert werden.
3. Die Mengenlehre und die damit verbundene Theorie der Relationen bietet die Grundlage der Theorie der relationalen Datenbanken.

Die obige Auflistung ließe sich leicht fortsetzen, aber neben den unmittelbaren Anwendungen von Logik und Mengenlehre hat die Beschäftigung mit diesen beiden Gebieten noch eine andere, sehr wichtige Funktion: Ohne die Einführung geeigneter Abstraktionen sind komplexe Systeme nicht beherrschbar. Kein Mensch ist in der Lage, alle Details eines Software-Systems, das aus mehreren 100 000 Programm-Zeilen besteht, zu verstehen. Die einzige Chance um ein solches System zu beherrschen besteht in der Einführung geeigneter Abstraktionen. Daher gehört ein überdurchschnittliches Abstraktionsvermögen zu den wichtigsten Werkzeugen eines Informatikers. Die Beschäftigung mit Logik und Mengenlehre trainiert gerade dieses abstrakte Denkvermögen.

Aus meiner Erfahrung weiß ich, dass einige der Studenten sich unter dem Thema Informatik etwas Anderes vorgestellt haben als die Diskussion abstrakter Konzepte. Für diese Studenten ist die Beherrschung einer Programmiersprache und einer dazugehörigen Programmierumgebung das Wesentliche der Informatik. Natürlich ist die Beherrschung einer Programmiersprache für einen Informatiker unabdingbar. Sie sollten sich allerdings darüber im klaren sein, dass das damit verbundene Wissen sehr vergänglich ist, denn niemand kann heute sagen, in welcher Programmiersprache in 10 Jahren programmiert werden wird. Im Gegensatz dazu sind die mathematischen Grundlagen der Informatik wesentlich beständiger.

1.2 Überblick über den Inhalt der Vorlesung

Die erste Informatik-Vorlesung legt die Grundlagen, die für das weitere Studium der Informatik benötigt werden. Bei diesen Grundlagen handelt es sich im wesentlichen um die *mathematische Logik*. Als Anwendung dieser Grundlagen werden zwei Programmiersprachen vorgestellt:

1. SETLX (set language extended) ist eine *mengenbasierte* Programmiersprache, in der dem Programmierer die Operationen der Mengenlehre zur Verfügung gestellt werden. Zusätzlich unterstützt die Sprache *funktionales Programmieren*.
2. *Prolog* (programming in logic) ist eine *deklarative* Programmier-Sprache, bei der die Programme die Form von logischen Formeln haben. Diese Sprache wird bei KI-Projekten und zum Prototyping eingesetzt.

Die Begriffs-Bildungen der Mengenlehre sind nicht sehr kompliziert, dafür aber um so abstrakter. Um diese Begriffs-Bildungen konkreter werden zu lassen und darüber hinaus den Studenten ein Gefühl für die Nützlichkeit

der Mengenlehre zu geben, stellen wir im dritten Kapitel die Sprache SETLX vor. Dies ist eine Programmiersprache, die auf der Mengenlehre aufgebaut ist. Dadurch ist es in SETLX möglich, Algorithmen in der Sprache der Mengenlehre zu formulieren. Die Formulierung eines Algorithmus in der Mengenlehre ist in der Regel wesentlich klarer als die Implementierung des Algorithmus in einer klassischen Programmiersprache wie beispielsweise C oder Java. Wir werden uns daher im zweiten Kapitel zunächst mit der Mengenlehre beschäftigen um dann im dritten Kapitel die Programmiersprache SETLX als Anwendung der Mengenlehre vorstellen zu können.

In dem vierten Kapitel widmen wir uns der *Aussagen-Logik*. Die Handhabung aussagenlogischer Formeln ist einfacher als die Handhabung prädikatenlogischer Formeln. Daher bietet sich die Aussagen-Logik gewissermaßen als Trainings-Objekt an um mit den Methoden der Logik vertraut zu werden. Die Aussagen-Logik hat gegenüber der Prädikaten-Logik noch einen weiteren Vorteil: Sie ist *entscheidbar*, d.h. wir können ein Programm schreiben, dass als Eingabe eine aussagenlogische Formel verarbeitet und welches dann entscheidet, ob diese Formel gültig ist. Ein solches Programm existiert für beliebige Formeln der Prädikaten-Logik nicht. Darüber hinaus gibt es in der Praxis eine Reihe von Problemen, die bereits mit Hilfe der Aussagenlogik gelöst werden können. Beispielsweise läßt sich die Frage nach der Korrektheit kombinatorischer digitaler Schaltungen auf die Entscheidbarkeit einer aussagenlogischen Formel zurückführen. Außerdem gibt es eine Reihe kombinatorischer Probleme, die sich auf aussagenlogische Probleme reduzieren lassen. Als ein Beispiel zeigen wir, wie sich das 8-Damen-Problem mit Hilfe der Aussagenlogik lösen läßt.

Im fünften Kapitel behandeln wir die Prädikatenlogik und analysieren den Begriff des prädikatenlogischen Beweises mit Hilfe eines *Kalküls*. Ein *Kalkül* ist dabei ein formales Verfahren, einen mathematischen Beweis zu führen. Ein solches Verfahren läßt sich programmieren. Wir stellen zu diesem Zweck den *Resolutions-Kalkül* vor, mit dem sich Beweise führen lassen. Dieser Resolutions-Kalkül bildet außerdem die Grundlage für die Programmiersprache *Prolog*, die wir im sechsten Kapitel diskutieren.

Bemerkung: Zum Schluß möchte ich hier noch ein Paar Worte zum Gebrauch von neuer und alter Rechtschreibung und der Verwendung von Spell-Checkern in diesem Skript sagen. Dieses Skript wurde unter Verwendung strengster wirtschaftlicher Kriterien erstellt. Im Klartext heißt das: Zeit ist Geld und als Dozent an der DHBW hat man weder das eine noch das andere. Daher ist es sehr wichtig zu wissen, wo eine zusätzliche Investition von Zeit noch einen für die Studenten nützlichen Effekt bringt und wo dies nicht der Fall ist. Ich habe mich an aktuellen Forschungs-Ergebnissen zum Nutzen der Rechtschreibung orientiert. Diese zeigen, daß es nicht wichtig ist, in welcher Reihenfolge die Buchstaben in einem Wort stehen, das einzige was wichtig ist, ist dass der erste und der letzte Buchstabe an der richtigen Position steht. Der Rest kann ein toller Böldisinn sein, trotzdem kann man ihn ohne Probleme lesen. Das ist so, weil wir nicht jeden Buchstaben einzeln lesen, sondern das Wort als Gesamtes. Wie sie sehen, ist das tatsächlich der Fall. ☺

Nichtsdestotrotz möchte ich Sie darum bitten, mir Tipp- und sonstige Fehler, die Ihnen in diesem Skript auffallen, per Email zu melden. Es bringt wenig, wenn Sie mir diese Fehler nach der Vorlesung sagen, denn bis ich dazu komme, die Fehler zu korrigieren, habe ich längst vergessen, was das Problem war.

Kapitel 2

Mathematische Grundlagen

In diesem Kapitel führen wir den Begriff der *prädikatenlogischen Formel* informell als Kurzschreibweise ein. Um später eine mathematisch einwandfreie Definition der *Semantik*, also der Bedeutung dieser Formeln geben zu können, benötigen wir die Mengenlehre, die wir im zweiten Abschnitt dieses Kapitels einführen. Anschließend diskutieren wir *binäre Relationen* und diskutieren *transitive Relationen* und *Äquivalenz-Relationen*.

2.1 Prädikatenlogische Formeln

Der Begriff der *prädikatenlogischen Formel* wird in dieser Vorlesung eine zentrale Rolle spielen. Wir werden prädikatenlogische Formeln zunächst als *Abkürzungen* definieren. Vorher motivieren wir die Verwendung solcher Formeln.

2.1.1 Warum Formeln

Betrachten wir einmal den folgenden mathematischen Text:

Addieren wir zwei Zahlen und bilden dann das Quadrat dieser Summe, so ist das Ergebnis das selbe, wie wenn wir zunächst beide Zahlen einzeln quadrieren, diese Quadrate aufsummieren und dazu noch das Produkt der beiden Zahlen zweifach hinzu addieren.

Der mathematische Satz, der hier ausgedrückt wird, ist Ihnen aus der Schule bekannt, es handelt sich um den ersten Binomischen Satz. Um dies zu sehen, führen wir für die in dem Text genannten zwei Zahlen die Variablen a und b ein und übersetzen dann die in dem obigen Text auftretenden Teilsätze in Terme. Die folgende Tabelle zeigt diesen Prozeß:

<i>Addieren wir zwei Zahlen</i>	$a + b$
<i>bilden das Quadrat dieser Summe</i>	$(a + b)^2$
<i>beide Zahlen einzeln quadrieren</i>	a^2, b^2
<i>diese Quadrate aufsummieren</i>	$a^2 + b^2$
<i>das Produkt der beiden Zahlen ...</i>	$a \cdot b$
<i>... zweifach hinzu addieren</i>	$a^2 + b^2 + 2 \cdot a \cdot b$

Insgesamt finden wir so, dass der obige Text zu der folgenden Formel äquivalent ist:

$$(a + b)^2 = a^2 + b^2 + 2 \cdot a \cdot b.$$

Für den mathematisch Geübten ist diese Formel offensichtlich leichter zu verstehen als der oben angegebene Text. Aber die Darstellung von mathematischen Zusammenhängen durch Formeln bietet neben der verbesserten Lesbarkeit noch zwei weitere Vorteile:

1. Formeln sind *manipulierbar*, d. h. wir können mit Formeln *rechnen*. Außerdem lassen Formeln sich aufgrund ihrer vergleichsweise einfachen Struktur auch mit Hilfe von Programmen bearbeiten und analysieren. Beim heutigen Stand der Technik ist es hingegen nicht möglich, natürlichsprachlichen Text mit dem Rechner vollständig zu analysieren und zu verstehen.
2. Darüber hinaus läßt sich die Bedeutung von Formeln mathematisch definieren und steht damit zweifelsfrei fest. Eine solche mathematische Definition der Bedeutung ist für natürlichsprachlichen Text so nicht möglich, da natürlichsprachlicher Text oft mehrdeutig ist und die genaue Bedeutung nur aus dem Zusammenhang hervorgeht. Als Beispiel für die Mehrdeutigkeit natürlicher Sprache möchte ich die folgende Schlagzeile aus der Zeitung *Journal Star* von 1980 präsentieren:

Crowd Rushing to See Pope Tramples 6 to Death

Diesen Satz können Sie auf zwei Arten lesen. Gemeint ist eigentlich, dass unter der Gruppe von Menschen, die den Papst sehen wollten, eine Panik ausbrach, was zum Tod von sechs Menschen führte. Die andere Interpretation dieses Satzes zeigt den Papst in einer aktiveren Rolle, in der er an die guten alten Zeiten der Heiligen Römischen Inquisition anknüpft.

2.1.2 Formeln als Kurzschreibweise

Nach dieser kurzen Motivation führen wir zunächst Formeln als Abkürzungen ein und stellen der Reihe nach die Ingredienzen vor, die wir zum Aufbau einer Formel benötigen.

1. Variablen

Variablen dienen uns als Namen für verschieden Objekte. Oben, bei dem Beispiel mit der ersten binomischen Formel, haben wir beispielsweise für die beiden zu addierenden Zahlen die Variablen a und b eingeführt. Die Idee bei der Einführung einer Variable ist, dass diese ein Objekt bezeichnet, dessen Identität noch nicht feststeht.

2. Konstanten

Konstanten bezeichnen Objekte, deren Identität feststeht. In der Mathematik werden beispielsweise Zahlen wie 1 oder π als Konstanten verwendet. Würden wir Aussagen über den biblischen Stammbaum als Formeln darstellen, so würden wir **Adam** und **Eva** als Konstanten verwenden.

Dieses letzte Beispiel mag Sie vielleicht verwundern, weil Sie davon ausgehen, dass Formeln nur dazu benutzt werden, mathematische oder allenfalls technische Zusammenhänge zu beschreiben. Der logische Apparat ist aber keineswegs auf eine Anwendung in diesen Bereichen beschränkt. Gerade auch Sachverhalte aus dem täglichen Leben lassen sich mit Hilfe von Formeln präzise beschreiben. Das ist auch notwendig, denn wir wollen ja später unsere Formeln zur Analyse von Programmen benutzen und diese Programme werden sich durchaus auch mit der Lösung von Problemen beschäftigen, die ihren Ursprung außerhalb der Technik haben.

Variablen und Konstanten werden zusammenfassend als *atomare Terme* bezeichnet. Das Attribut *atomar* bezieht sich hierbei auf die Tatsache, dass diese Terme sich nicht weiter in Bestandteile zerlegen lassen. Im Gegensatz dazu stehen die *zusammengesetzten Terme*. Dies sind Terme, die mit Hilfe von Funktions-Zeichen aus anderen Termen aufgebaut werden.

3. Funktions-Zeichen

Funktions-Zeichen benutzen wir, um aus Variablen und Konstanten neue Ausdrücke aufzubauen, die wiederum Objekte bezeichnen. In dem obigen Beispiel haben wir das Funktions-Zeichen “+” benutzt und mit diesem Funktions-Zeichen aus den Variablen a und b den Ausdruck $a + b$ gebildet. Allgemein nennen wir Ausdrücke, die sich aus Variablen, Konstanten und Funktions-Zeichen bilden lassen, *Terme*.

Das Funktions-Zeichen “+” ist zweistellig, aber natürlich gibt es auch einstellige und mehrstellige Funktions-Zeichen. Ein Beispiel aus der Mathematik für ein einstelliges Funktions-Zeichen ist das Zeichen “ $\sqrt{}$ ”. Ein weiteres Beispiel ist durch das Zeichen “sin” gegeben, dass in der Mathematik für die Sinus-Funktion verwendet wird.

Allgemein gilt: Ist f ein n -stelliges Funktions-Zeichen und sind t_1, \dots, t_n Terme, so kann mit Hilfe des Funktions-Zeichen f daraus der neue Term

$$f(t_1, \dots, t_n)$$

gebildet werden. Diese Schreibweise, bei der zunächst das Funktions-Zeichen gefolgt von einer öffnenden Klammer angegeben wird und anschließend die Argumente der Funktion durch Kommata getrennt aufgelistet werden, gefolgt von einer schließenden Klammer, ist der “Normalfall”. Diese Notation wird auch als *Prefix-Notation* bezeichnet. Bei einigen zweistelligen Funktions-Zeichen hat es sich aber eingebürgert, diese in einer *Infix-Notation* darzustellen, d.h. solche Funktions-Zeichen werden zwischen die Terme geschrieben. In der Mathematik liefern die Funktions-Zeichen “+”, “−”, “.” und “/” hierfür Beispiele. Schließlich gibt es noch Funktions-Zeichen, die auf ihr Argument folgen. Ein Beispiel dafür ist das Zeichen “!” zur Bezeichnung der *Fakultät*¹, denn für die Fakultät einer Zahl n hat sich in der Mathematik die Schreibweise “ $n!$ ” eingebürgert. Eine solche Notation wird als *Postfix-Notation* bezeichnet.

4. Prädikate

Prädikate stellen zwischen verschiedenen Objekten eine Beziehung her. Ein wichtiges Prädikat ist das Gleichheits-Prädikat, dass durch das Gleichheits-Zeichen “=” dargestellt wird. Setzen wir zwei Terme t_1 und t_2 durch das Gleichheits-Zeichen in Beziehung, so erhalten wir die *Formel* $t_1 = t_2$.

Genau wie Funktions-Zeichen auch hat jedes Prädikat eine vorgegebene *Stelligkeit*. Diese gibt an, wie viele Objekte durch das Prädikat in Relation gesetzt werden. Im Falle des Gleichheits-Zeichens ist die Stelligkeit 2, aber es gibt auch Prädikate mit anderen Stelligkeiten. Zum Beispiel könnten wir ein Prädikat “istQuadrat” definieren, dass für natürliche Zahlen ausdrückt, dass diese Zahl eine Quadrat-Zahl ist. Ein solches Prädikat wäre dann einstellig.

Ist allgemein p ein n -stelliges Prädikats-Zeichen und sind die Ausdrücke t_1, \dots, t_n Terme, so kann aus diesen Bestandteilen die *Formel*

$$p(t_1, \dots, t_n)$$

gebildet werden. Formeln von dieser Bauart bezeichnen wir auch als *atomare Formel*, denn sie ist zwar aus Termen, nicht jedoch aus anderen Formeln zusammengesetzt.

Genau wie bei zweistelligen Funktions-Zeichen hat sich auch bei zweistelligen Prädikats-Zeichen eine *Infix-Notation* eingebürgert. Das Prädikats-Zeichen “=” liefert ein Beispiel hierfür, denn wir schreiben “ $a = b$ ” statt “ $= (a, b)$ ”. Andere Prädikats-Zeichen, für die sich eine Infix-Notation eingebürgert hat, sind die Prädikats-Zeichen “<”, “≤”, “>” und “≥”, die zum Vergleich von Zahlen benutzt werden.

Bemerkung: Der Aufbau einer atomaren Formel $p(t_1, \dots, t_m)$ ist ganz ähnlich zu dem Aufbau eines Terms der Form $f(t_1, \dots, t_m)$, aber die Bedeutung ist eine ganz Andere: Die Formel $p(t_1, \dots, t_m)$ ist entweder wahr oder falsch, während der Term $f(t_1, \dots, t_n)$ ein Objekt bezeichnet. Um atomare Formeln und Terme unterscheiden zu können müssen wir Prädikats-Zeichen und Funktions-Zeichen unterscheiden können.

5. Junktoren

Junktoren werden dazu benutzt, Formeln mit einander in Beziehung zu setzen. Der einfachste Junktor ist das “und”. Haben wir zwei Formeln F_1 und F_2 und wollen ausdrücken, dass sowohl F_1 als auch F_2 gültig ist, so schreiben wir

$$F_1 \wedge F_2$$

und lesen dies als “ F_1 und F_2 ”. Die nachfolgende Tabelle listet alle Junktoren auf, die wir verwenden werden:

¹ Für eine positive natürliche Zahl n ist die *Fakultät* von n als das Produkt aller natürlichen Zahlen von 1 bis n definiert. Die Fakultät von n wird mit $n!$ bezeichnet, es gilt also $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-1) \cdot n$.

Junktor	Bedeutung
$\neg F$	nicht F
$F_1 \wedge F_2$	F_1 und F_2
$F_1 \vee F_2$	F_1 oder F_2
$F_1 \rightarrow F_2$	wenn F_1 , dann F_2
$F_1 \leftrightarrow F_2$	F_1 genau dann, wenn F_2

Hier ist noch zu bemerken, dass es bei komplexeren Formeln zur Vermeidung von Mehrdeutigkeiten notwendig ist, diese geeignet zu klammern. Bezeichnen beispielsweise P , Q und R atomare Formeln, so können wir unter Zuhilfenahme von Klammern daraus die folgenden Formeln bilden:

$$P \rightarrow (Q \vee R) \quad \text{und} \quad (P \rightarrow Q) \vee R.$$

Umgangssprachlich würden beide Formeln wie folgt interpretiert:

Aus P folgt Q oder R .

Die mathematische Schreibweise ist hier durch die Klammersetzung eindeutig, während die umgangssprachlichen Formulierung zwei verschiedene Interpretationen zulässt.

Die Verwendung von vielen Klammern vermindert die Lesbarkeit einer Formel. Um Klammern einsparen zu können, vereinbaren wir daher ähnliche Bindungsregeln, wie wir sie aus der Schulmathematik kennen. Dort wurde vereinbart, dass “+” und “−” schwächer binden als “·” und “/” und damit ist gemeint, dass

$$x + y \cdot z \quad \text{als} \quad x + (y \cdot z)$$

interpretiert wird. Ähnlich vereinbaren wir hier, dass “ \neg ” stärker bindet als “ \wedge ” und “ \vee ” und dass diese beiden Operatoren stärker binden als “ \rightarrow ”. Schließlich bindet der Operator “ \leftrightarrow ” schwächer als alle anderen Operatoren. Mit diesen Vereinbarungen lautet die Formel

$$P \wedge Q \rightarrow R \leftrightarrow \neg R \rightarrow \neg P \vee \neg Q$$

dann in einer vollständig geklammerten Schreibweise

$$((P \wedge Q) \rightarrow R) \leftrightarrow ((\neg R) \rightarrow ((\neg P) \vee (\neg Q))).$$

Bemerkung: In der Literatur wird gelegentlich gefordert, dass der Operator “ \wedge ” stärker bindet als der Operator “ \vee ” und dieser Konvention folgt auch die Programmiersprache **C**, in der das logische **Und** durch den Operator “ $\&\&$ ” und das logische **Oder** durch “ $||$ ” dargestellt wird. Eine solche Konvention ist allerdings später, wenn wir Formeln in die sogenannte *konjunktive Normalform* bringen, kontraintuitiv. Daher haben in diesem Skript die beiden Operatoren “ \wedge ” und “ \vee ” die selbe Bindungsstärke.

Bemerkung: Die Bindungsstärke eines Operators wird auch als die *Präzedenz* des Operators bezeichnet.

6. *Quantoren* geben an, in welcher Weise eine Variable in einer Formel interpretiert wird. Wir führen zwei Quantoren ein, den All-Quantor “ \forall ” und den Existenz-Quantor “ \exists ”. Eine Formel der Form

$$\forall x : F$$

lesen wir als “für alle x gilt F ” und eine Formel der Form

$$\exists x : F$$

wird als “es gibt ein x , so dass F gilt” gelesen. In dieser Vorlesung werden wir üblicherweise *qualifizierte Quantoren* verwenden. Die Qualifizierung gibt dabei an, in welchem Bereich die durch die Variablen bezeichneten Objekte liegen müssen. Im Falle des All-Quantors schreiben wir dann

$$\forall x \in M : F$$

und lesen dies als “für alle x aus M gilt F ”. Hierbei bezeichnet M eine Menge. Dies ist nur eine abkürzende Schreibweise, die wir wie folgt definieren können:

$$\forall x \in M : F \stackrel{\text{def}}{\iff} \forall x : (x \in M \rightarrow F)$$

Entsprechend lautet die Notation für den Existenz-Quantor

$$\exists x \in M : F$$

und das wird dann als “*es gibt ein x aus M , so dass F gilt*” gelesen. Formal läßt sich das als

$$\exists x \in M : F \stackrel{\text{def}}{\iff} \exists x : (x \in M \wedge F)$$

definieren. Wir verdeutlichen die Schreibweisen durch ein Beispiel. Die Formel

$$\forall x \in \mathbb{R} : \exists n \in \mathbb{N} : n > x$$

lesen wir wie folgt:

Für alle x aus \mathbb{R} gilt: Es gibt ein n aus \mathbb{N} , so dass n größer als x ist.

Hier steht \mathbb{R} für die reellen Zahlen und \mathbb{N} bezeichnet die natürlichen Zahlen. Die obige Formel drückt also aus, dass es zu jeder reellen Zahl x eine natürliche Zahl n gibt, so dass n größer als x ist.

Treten in einer Formel Quantoren und Junktoren gemischt auf, so stellt sich die Frage, was stärker bindet. Wir vereinbaren, dass Quantoren schwächer binden als Junktoren. Eine Formel der Art

$$\forall x : p(x) \wedge q(x)$$

wird damit implizit wie folgt geklammert:

$$\forall x : (p(x) \wedge q(x)).$$

2.1.3 Beispiele für Terme und Formeln

Um die Konzepte “Term” und “Formel” zu verdeutlichen, geben wir im folgenden einige Beispiele an. Wir wählen ein Beispiel aus dem täglichen Leben und geben Terme und Formeln an, die sich mit Verwandschaftsbeziehungen beschäftigen. Wir beginnen damit, dass wir die Konstanten, Variablen, Funktions-Zeichen und Prädikats-Zeichen festlegen.

1. Als *Konstanten* verwenden wir die Wörter

“adam”, “eva”, “kain” und “abel”, “lisa”.

2. Als *Variablen* verwenden wir die Buchstaben

“ x ”, “ y ” und “ z ”.

3. Als *Funktions-Zeichen* verwenden wir die Wörter

“vater” und “mutter”.

Diese beiden Funktions-Zeichen sind einstellig.

4. Als *Prädikats-Zeichen* verwenden wir die Wörter

“bruder”, “schwester”, “onkel”, “männlich” und “weiblich”.

Alle diese Prädikats-Zeichen sind zweistellig. Als weiteres zweistelliges Prädikats-Zeichen verwenden wir das Gleichheits-Zeichen “=”.

Eine solche Ansammlung von Konstanten, Variablen, Funktions-Zeichen und Prädikats-Zeichen bezeichnen wir auch als *Signatur*. Wir geben zunächst einige Terme an, die sich mit dieser Signatur bilden lassen:

1. “kain” ist ein Term, denn “kain” ist eine Konstante.
2. “vater(kain)” ist ein Term, denn “kain” ist ein Term und “vater” ist ein einstelliges Funktions-Zeichen.

3. “**mutter**(**vater**(**kain**))” ist ein Term, denn “**vater**(**kain**)” ist ein Term und “**mutter**” ist ein einstelliges Funktions-Zeichen,
4. “**männlich**(**kain**)” ist eine Formel, denn “**kain**” ist ein Term und “**männlich**” ist ein einstelliges Prädikats-Zeichen.
5. “**männlich**(**lisa**)” ist ebenfalls eine Formel, denn “**lisa**” ist ein Term.

Dieses Beispiel zeigt, dass Formeln durchaus auch falsch sein können. Die bisher gezeigten Formeln sind alle atomar. Wir geben nun Beispiele für zusammengesetzte Formeln.

6. “**vater**(x) = **vater**(y) \wedge **mutter**(x) = **mutter**(y) \rightarrow **bruder**(x, y) \vee **schwester**(x, y)”
ist eine Formel, die aus den beiden Formeln

$$\begin{aligned} &\text{“vater}(x) = \text{vater}(y) \wedge \text{mutter}(x) = \text{mutter}(y)\text{”} \quad \text{und} \\ &\text{“bruder}(x, y) \vee \text{schwester}(x, y)\text{”} \end{aligned}$$

aufgebaut ist.

7. “ $\forall x: \forall y: \text{bruder}(x, y) \vee \text{schwester}(x, y)$ ” ist eine Formel.

Die Formel Nr. 7 ist intuitiv gesehen falsch. Auch die Formel Nr. 6 ist falsch, wenn wir davon ausgehen, dass niemand sein eigener Bruder ist. Um die Begriffe “*wahr*” und “*falsch*” für Formeln streng definieren zu können, ist es notwendig, die *Interpretation* der verwendeten Signatur festzulegen. Anschaulich gesehen definiert eine *Interpretation* die Bedeutung der *Symbole*, also der Konstanten, Funktions- und Prädikats-Zeichen, aus denen die Signatur besteht. Exakt kann der Begriff aber erst angegeben werden, wenn Hilfsmittel aus der Mengenlehre zur Verfügung stehen. Dieser wenden wir uns jetzt zu.

2.2 Mengen und Relationen

Die Mengenlehre ist gegen Ende des 19-ten Jahrhunderts aus dem Bestreben heraus entstanden, die Mathematik auf eine solide Grundlage zu stellen. Die Schaffung einer solchen Grundlage wurde als notwendig erachtet, da der Begriff der Unendlichkeit den Mathematikern zunehmend Kopfzerbrechen bereitete.

Begründet wurde die Mengenlehre in wesentlichen Teilen von Georg Cantor (1845 – 1918). Die erste Definition des Begriffs der Menge lautete etwa wie folgt [Can95]:

Eine *Menge* ist eine *wohldefinierte* Ansammlung unterscheidbarer Objekte zu einem Ganzen.

Das Attribut “*wohldefiniert*” drückt dabei aus, dass wir für eine vorgegebene Menge M und ein Objekt x stets klar sein muss, ob das Objekt x zu der Menge M gehört oder nicht. In diesem Fall schreiben wir

$$x \in M$$

und lesen diese Formel als “ x ist ein *Element* der Menge M ”. Das Zeichen “ \in ” wird in der Mengenlehre also als zweistelliges Prädikats-Zeichen gebraucht, für das sich eine Infix-Notation eingebürgert hat. Um den Begriff der Menge mathematisch zu präzisieren, führte Cantor das sogenannte *Komprehensions-Axiom* ein. Wir können dieses zunächst wie folgt formalisieren: Ist $p(x)$ eine Eigenschaft, die ein Objekt x entweder hat oder nicht, so können wir die Menge M aller Objekte, welche die Eigenschaft $p(x)$ haben, bilden. Wie schreiben dann

$$M = \{x \mid p(x)\}$$

und lesen dies als “ M ist die Menge aller x , auf welche die Eigenschaft $p(x)$ zutrifft”. Eine Eigenschaft $p(x)$ ist dabei nichts anderes als eine Formel, in der die Variable x vorkommt. Wir veranschaulichen das Komprehensions-Axiom durch ein Beispiel: Es sei \mathbb{N} die Menge der natürlichen Zahlen. Ausgehend von der Menge \mathbb{N} wollen wir die Menge der *geraden Zahlen* definieren. Zunächst müssen wir dazu die Eigenschaft einer Zahl x , *gerade* zu sein, durch eine Formel $p(x)$ mathematisch erfassen. Eine natürliche Zahl x ist genau dann gerade, wenn es eine natürliche Zahl y gibt, so dass x das Doppelte von y ist. Damit können wir die Eigenschaft $p(x)$ folgendermaßen definieren:

$$p(x) := (\exists y \in \mathbb{N} : x = 2 \cdot y).$$

Also kann die Menge der geraden Zahlen als

$$\{x \mid \exists y \in \mathbb{N} : x = 2 \cdot y\}$$

geschrieben werden.

Leider führt die uneingeschränkte Anwendung des Komprehensions-Axiom schnell zu Problemen. Betrachten wir dazu die Eigenschaft einer Menge, sich nicht selbst zu enthalten, wir setzen also $p(x) := \neg x \in x$ und definieren die Menge R als

$$R := \{x \mid \neg x \in x\}.$$

Intuitiv würden wir vielleicht erwarten, dass keine Menge sich selbst enthält. Wir wollen jetzt zunächst für die eben definierte Menge R überprüfen, wie die Dinge liegen. Es können zwei Fälle auftreten:

1. Fall: $\neg R \in R$. Also enthält die Menge R sich nicht selbst. Da die Menge R aber als die Menge der Mengen definiert ist, die sich nicht selber enthalten, müßte R eine Element von R sein, es müßte also $R \in R$ gelten im Widerspruch zur Voraussetzung $\neg R \in R$.
2. Fall: $R \in R$. Setzen wir hier die Definition von R ein, so haben wir

$$R \in \{x \mid \neg x \in x\}.$$

Dass heißt dann aber gerade $\neg R \in R$ und steht im Widerspruch zur Voraussetzung $R \in R$.

Wie wir es auch drehen und wenden, es kann weder $R \in R$ noch $\neg R \in R$ gelten. Als Ausweg können wir nur feststellen, dass das vermittels

$$\{x \mid \neg x \in x\}$$

definierte Objekt keine Menge ist. Das heißt dann aber, dass das Komprehensions-Axiom zu allgemein ist. Wir folgern, dass nicht jede in der Form

$$M = \{x \mid p(x)\}$$

angegebene Menge existiert. Die Konstruktion der "Menge" $\{x \mid \neg x \in x\}$ stammt von dem britischen Logiker und Philosophen Bertrand Russell (1872 – 1970). Sie wird deswegen auch als *Russell'sche Antinomie* bezeichnet.

Um Paradoxien wie die Russell'sche Antinomie zu vermeiden, ist es erforderlich, bei der Konstruktion von Mengen vorsichtiger vorzugehen. Wir werden im folgenden Konstruktions-Prinzipien für Mengen vorstellen, die schwächer sind als das Komprehensions-Axiom, die aber für die Praxis der Informatik ausreichend sind. Wir wollen dabei die dem Komprehensions-Axiom zugrunde liegende Notation beibehalten und Mengendefinitionen in der Form

$$M = \{x \mid p(x)\}$$

angeben. Um Paradoxien zu vermeiden, werden wir nur bestimmte Sonderfälle dieser Gleichung zulassen. Diese Sonderfälle, sowie weitere Möglichkeiten Mengen zu konstruieren, stellen wir jetzt vor.

2.2.1 Erzeugung von Mengen durch explizites Auflisten

Die einfachste Möglichkeit, eine Menge festzulegen, besteht in der expliziten *Auflistung* aller ihrer Elemente. Diese Elemente werden in den geschweiften Klammern "{" und "}" eingefaßt und durch Kommas getrennt. Definieren wir beispielsweise

$$M := \{1, 2, 3\},$$

so haben wir damit festgelegt, dass die Menge M aus den Elementen 1, 2 und 3 besteht. In der Schreibweise des Komprehensions-Axioms können wir diese Menge als

$$M = \{x \mid x = 1 \vee x = 2 \vee x = 3\}$$

angeben.

Als ein weiteres Beispiel für eine Menge, die durch explizite Aufzählung ihrer Elemente angegeben werden kann, betrachten wir die Menge B der kleinen lateinischen Buchstaben, die wir als

$$B := \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}.$$

definieren können.

2.2.2 Die Menge der natürlichen Zahlen

Alle durch explizite Auflistung definierten Mengen haben offensichtlich nur endlich viele Elemente. Aus der mathematischen Praxis kennen wir aber auch Mengen mit unendlich vielen Elementen. Ein Beispiel ist die Menge der natürlichen Zahlen, die wir mit \mathbb{N} bezeichnen. Mit den bisher behandelten Verfahren läßt diese Menge sich nicht definieren. Wir müssen daher die Existenz dieser Menge als Axiom fordern. Intuitiv schreiben wir

$$\mathbb{N} := \{0, 1, 2, 3, \dots\}.$$

Eine exakte Definition ist das aber nicht. Die Menge \mathbb{N} der natürlichen Zahlen kann exakt wie folgt definiert werden: Die Menge \mathbb{N} der natürlichen Zahlen ist die kleinste Menge, für die Folgendes gilt:

1. $0 \in \mathbb{N}$,
2. $\forall n : n \in \mathbb{N} \rightarrow (n + 1) \in \mathbb{N}$.

Eine Definition der obigen Form wird als *induktive* Definition bezeichnet.

Neben der Menge \mathbb{N} der natürlichen Zahlen verwenden wir noch die folgenden Mengen von Zahlen:

1. \mathbb{Z} : Menge der ganzen Zahlen.
2. \mathbb{Q} : Menge der rationalen Zahlen.
3. \mathbb{R} : Menge der reellen Zahlen.

In der Mathematik wird gezeigt, wie sich diese Mengen aus der Menge der natürlichen Zahlen erzeugen lassen.

2.2.3 Das Auswahl-Prinzip

Das *Auswahl-Prinzip* ist eine Abschwächung des Komprehensions-Axiom. Die Idee ist, mit Hilfe einer Eigenschaft p aus einer schon vorhandenen Menge M die Menge N der Elemente x *auszuwählen*, die eine bestimmte Eigenschaft $p(x)$ besitzen:

$$N = \{x \in M \mid p(x)\}$$

In der Notation des Komprehensions-Axioms schreibt sich diese Menge als

$$N = \{x \mid x \in M \wedge p(x)\}.$$

Im Unterschied zu dem Komprehensions-Axiom können wir uns hier nur auf die Elemente einer bereits vorgegebenen Menge M beziehen und nicht auf völlig beliebige Objekte.

Beispiel: Die Menge der geraden Zahlen kann mit dem Auswahl-Prinzip wie folgt definiert werden:

$$\{x \in \mathbb{N} \mid \exists y \in \mathbb{N} : x = 2 \cdot y\}.$$

◇

2.2.4 Vereinigungs-Mengen

Sind zwei Mengen M und N gegeben, so enthält die Vereinigung von M und N alle Elemente, die in der Menge M oder in der Menge N liegen. Für diese Vereinigung schreiben wir $M \cup N$. Formal kann die Vereinigung wie folgt definiert werden:

$$M \cup N := \{x \mid x \in M \vee x \in N\}.$$

Beispiel: Ist $M = \{1, 2, 3\}$ und $N = \{2, 5\}$, so gilt:

$$\{1, 2, 3\} \cup \{2, 5\} = \{1, 2, 3, 5\}.$$

◇

Der Begriff der Vereinigung von Mengen läßt sich verallgemeinern. Betrachten wir dazu eine Menge X , deren Elemente selbst wieder Mengen sind. Beispielsweise ist die später betrachtete Potenz-Menge einer Menge von dieser Art. Wir können dann die Vereinigung aller Mengen, die Elemente von der Menge X sind, bilden. Diese Vereinigung schreiben wir als $\bigcup X$. Formal kann das wie folgt definiert werden:

$$\bigcup X = \{y \mid \exists x \in X : y \in x\}.$$

Beispiel: Die Menge X sei wie folgt gegeben:

$$X = \{\{\}, \{1, 2\}, \{1, 3, 5\}, \{7, 4\}\}.$$

Dann gilt

$$\bigcup X = \{1, 2, 3, 4, 5, 7\}.$$

◇

2.2.5 Schnitt-Menge

Sind zwei Mengen M und N gegeben, so definieren wir den *Schnitt* von M und N als die Menge aller Elemente, die sowohl in M als auch in N auftreten. Wir bezeichnen den Schnitt von M und N mit $M \cap N$. Formal können wir $M \cap N$ wie folgt definieren:

$$M \cap N := \{x \mid x \in M \wedge x \in N\}.$$

Beispiel: Wir berechnen den Schnitt der Mengen $M = \{1, 3, 5\}$ und $N = \{2, 3, 5, 6\}$. Es gilt

$$M \cap N = \{3, 5\}.$$

◇

Der Begriff der Schnitt-Menge von zwei Mengen läßt sich verallgemeinern. Betrachten wir dazu eine Menge X , deren Elemente selbst wieder Mengen sind. Wir können dann den Schnitt aller Mengen, die Elemente von der Menge X sind, bilden. Diese Vereinigung schreiben wir als $\bigcap X$. Formal kann das wie folgt definiert werden:

$$\bigcap X = \{y \mid \forall x \in X : y \in x\}.$$

Beispiel: Die Menge X sei gegeben als

$$X = \{\{1, 2, 3\}, \{2, 3, 4\}, \{1, 3, 5\}\}.$$

Dann gilt $\bigcap X = \{3\}$.

◇

2.2.6 Differenz-Mengen

Sind zwei Mengen M und N gegeben, so bezeichnen wir die *Differenz* von M und N als die Menge aller Elemente, die in M aber nicht in N auftreten. Wir schreiben hierfür $M \setminus N$. Das wird als *M ohne N* gelesen und kann formal wie folgt definiert werden:

$$M \setminus N := \{x \mid x \in M \wedge x \notin N\}.$$

Hier haben wir die Notation $x \notin N$ als Kurzschreibweise für die Formel $\neg x \in N$ verwendet.

Beispiel: Wir berechnen die Differenz der Mengen $M = \{1, 3, 5, 7\}$ und $N = \{2, 3, 5, 6\}$. Es gilt

$$M \setminus N = \{1, 7\}.$$

◇

2.2.7 Potenz-Mengen

Um den Begriff der *Potenz-Menge* einführen zu können, müssen wir zunächst *Teilmengen* definieren. Sind M und N zwei Mengen, so heißt M eine *Teilmenge* von N genau dann, wenn jedes Element der Menge M auch ein Element der Menge N ist. In diesem Fall schreiben wir $M \subseteq N$. Formal können wir den Begriff der Teilmenge also wie folgt einführen:

$$M \subseteq N \stackrel{\text{def}}{\iff} \forall x : (x \in M \rightarrow x \in N)$$

Unter der *Potenz-Menge* einer Menge M wollen wir nun die Menge aller Teilmengen von M verstehen. Wir schreiben 2^M für die Potenz-Menge von M . Dann gilt:

$$2^M = \{x \mid x \subseteq M\}.$$

Beispiel: Wir bilden die Potenz-Menge der Menge $\{1, 2, 3\}$. Es gilt:

$$2^{\{1,2,3\}} = \{\{\}, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Diese Menge hat $8 = 2^3$ Elemente. Allgemein kann durch Induktion nach der Anzahl n der Elemente der Menge M gezeigt werden, dass die Potenz-Menge einer Menge M , die aus n verschiedenen Elementen besteht, insgesamt 2^n Elemente enthält. Bezeichnen wir die Anzahl der Elemente einer endlichen Menge mit $\text{card}(M)$, so gilt also

$$\text{card}(2^M) = 2^{\text{card}(M)}. \quad (2.1)$$

Dies erklärt die Schreibweise 2^M für die Potenz-Menge von M . Wir beweisen die Gültigkeit der Gleichung 2.1 durch vollständige Induktion nach $n = \text{card}(M)$:

I.A.: $n = 0$

Aus $n = \text{card}(M) = 0$ folgt, dass die Menge M leer sein muss: $M = \{\}$. Damit haben wir

$$\text{card}(2^{\{\}}) = \text{card}(\{\{\}\}) = 1 = 2^0 = 2^{\text{card}(\{\})}$$

und damit ist die Behauptung im Fall $n = 0$ nachgewiesen.

I.S.: $n \mapsto n + 1$

Falls die Menge M aus $n + 1$ Elementen besteht, können wir M in der Form

$$M = A \cup \{b\}$$

schreiben, wobei A eine Menge mit n Elementen ist und b das $(n + 1)$ -te Element von M ist und damit das Element aus M ist, dass nicht bereits in der Teilmenge A enthalten ist. Für eine Teilmenge $C \subseteq M$ gibt es nun zwei sich ausschließende Möglichkeiten:

(a) $b \notin C$.

In diesem Fall ist C bereits eine Teilmenge von A .

(b) $b \in C$.

Dann ist C keine Teilmenge von A , aber C lässt sich dann in der Form

$$C = D \cup \{b\} \quad \text{mit } D \subseteq A$$

schreiben.

Insgesamt können wir daher die Potenz-Menge 2^M in zwei Teile zerlegen:

$$2^M = 2^A \cup \{D \cup \{b\} \mid D \in 2^A\}.$$

Diese beiden Teilmengen haben offenbar gleich viele Elemente und daher gilt

$$\begin{aligned}
 \text{card}(2^M) &= \text{card}(2^A) \cup \text{card}(\{D \cup \{b\} \mid D \in 2^A\}) \\
 &= \text{card}(2^A) + \text{card}(2^A) \\
 &= 2 \cdot \text{card}(2^A) \\
 &\stackrel{\text{IV}}{=} 2 \cdot 2^{\text{card}(A)} \\
 &= 2 \cdot 2^n \\
 &= 2^{n+1} \\
 &= 2^{\text{card}(M)}.
 \end{aligned}$$

Hier haben wir bei der mit IV markierten Gleichung ausgenutzt, dass wir nach Induktions-Voraussetzung bereits wissen, dass

$$\text{card}(2^A) = 2^{\text{card}(A)}$$

gilt, denn die Menge A enthält ja nur n Elemente. □

2.2.8 Bild-Mengen

Es sei M eine Menge und f sei eine Funktion, die für alle x aus M definiert ist. Dann heißt die Menge aller Abbilder $f(x)$ von Elementen x aus der Menge M das *Bild* von M unter f . Wir schreiben $f(M)$ für dieses Bild. Formal kann $f(M)$ wie folgt definiert werden:

$$f(M) := \{y \mid \exists x \in M : y = f(x)\}.$$

In der Literatur findet sich für die obige Menge auch die Schreibweise

$$f(M) = \{f(x) \mid x \in M\}.$$

Beispiel: Die Menge Q aller Quadrat-Zahlen kann wie folgt definiert werden:

$$Q := \{y \mid \exists x \in \mathbb{N} : y = x^2\}.$$

Alternativ können wir auch schreiben

$$Q = \{x^2 \mid x \in \mathbb{N}\}. \quad \diamond$$

2.2.9 Kartesische Produkte

Zur Einführung von Relationen und kartesischen Produkten benötigen wir den Begriff des *geordneten Paares* zweier Objekte x und y . Dieses wird als

$$\langle x, y \rangle$$

geschrieben. Wir sagen, dass x die *erste Komponente* des Paares $\langle x, y \rangle$ ist, und y ist die *zweite Komponente*. Zwei geordnete Paare $\langle x_1, y_1 \rangle$ und $\langle x_2, y_2 \rangle$ sind genau dann gleich, wenn sie komponentenweise gleich sind, d.h. es gilt

$$\langle x_1, y_1 \rangle = \langle x_2, y_2 \rangle \Leftrightarrow x_1 = x_2 \wedge y_1 = y_2.$$

Das kartesische Produkt zweier Mengen M und N ist nun die Menge aller geordneten Paare, deren erste Komponente in M liegt und deren zweite Komponente in N liegt. Das kartesische Produkt von M und N wird als $M \times N$ geschrieben, formal gilt:

$$M \times N := \{z \mid \exists x: \exists y: z = \langle x, y \rangle \wedge x \in M \wedge y \in N\}.$$

Alternativ können wir auch schreiben

$$M \times N := \{\langle x, y \rangle \mid x \in M \wedge y \in N\}.$$

Beispiel: Wir setzen $M = \{1, 2, 3\}$ und $N = \{5, 7\}$. Dann gilt

$$M \times N = \{\langle 1, 5 \rangle, \langle 2, 5 \rangle, \langle 3, 5 \rangle, \langle 1, 7 \rangle, \langle 2, 7 \rangle, \langle 3, 7 \rangle\}. \quad \diamond$$

Bemerkung: Sind M und N endliche Mengen, so gilt für die Anzahl der Elemente des kartesischen Produkts $M \times N$ die Beziehung

$$\text{card}(M \times N) = \text{card}(M) \cdot \text{card}(N).$$

Dieser Umstand erklärt, warum wir hier von einem *Produkt* sprechen. \diamond

Der Begriff des geordneten Paares läßt sich leicht zum Begriff des n -Tupels verallgemeinern: Ein n -Tupel hat die Form

$$\langle x_1, x_2, \dots, x_n \rangle.$$

Analog kann auch der Begriff des kartesischen Produktes auf n Mengen M_1, \dots, M_n verallgemeinert werden. Das sieht dann so aus:

$$M_1 \times \dots \times M_n = \{z \mid \exists x_1 \dots \exists x_n: z = \langle x_1, x_2, \dots, x_n \rangle \wedge x_1 \in M_1 \wedge \dots \wedge x_n \in M_n\}.$$

Ist f eine Funktion, die auf $M_1 \times \dots \times M_n$ definiert ist, so vereinbaren wir folgende Vereinfachung der Schreibweise:

$$f(x_1, \dots, x_n) \quad \text{steht für} \quad f(\langle x_1, \dots, x_n \rangle).$$

Gelegentlich werden n -Tupel auch als *endliche Folgen* oder als *Listen* bezeichnet.

2.2.10 Gleichheit von Mengen

Wir haben nun alle Verfahren, die wir zur Konstruktion von Mengen benötigen, vorgestellt. Wir klären jetzt die Frage, wann zwei Mengen gleich sind. Dazu postulieren wir das folgende *Extensionalitäts-Axiom* für Mengen:

Zwei Mengen sind genau dann gleich, wenn sie die selben Elemente besitzen.

Mathematisch können wir diesen Sachverhalt wie folgt ausdrücken:

$$M = N \leftrightarrow \forall x : (x \in M \leftrightarrow x \in N)$$

Eine wichtige Konsequenz aus diesem Axiom ist die Tatsache, dass die Reihenfolge, mit der Elemente in einer Menge aufgelistet werden, keine Rolle spielt. Beispielsweise gilt

$$\{1, 2, 3\} = \{3, 2, 1\},$$

denn beide Mengen enthalten die selben Elemente. Genausowenig spielt es eine Rolle, wie oft ein Element x in einer Menge vorkommt: Beispielsweise haben wir

$$\{3, 1, 3, 2, 3\} = \{1, 2, 3\},$$

denn beide Mengen enthalten die gleichen Elemente.

Falls Mengen durch explizite Aufzählung ihrer Elemente definiert sind, ist die Frage nach der Gleichheit zweier Mengen trivial. Ist eine der Mengen mit Hilfe des Auswahl-Prinzips definiert, so kann es beliebig schwierig sein zu entscheiden, ob zwei Mengen gleich sind. Hierzu ein Beispiel: Es läßt sich zeigen, dass

$$\{n \in \mathbb{N} \mid \exists x, y, z \in \mathbb{N} : x > 0 \wedge y > 0 \wedge x^n + y^n = z^n\} = \{1, 2\}$$

gilt. Allerdings ist der Nachweis dieser Gleichheit sehr schwer, denn er ist äquivalent zum Beweis der *Fermat'schen Vermutung*. Diese Vermutung wurde 1637 von *Pierre de Fermat* aufgestellt und konnte erst 1995 von Andrew Wiles bewiesen werden. Es gibt andere, ähnlich aufgebaute Mengen, wo bis heute unklar ist, welche Elemente in der Menge liegen und welche nicht.

2.2.11 Rechenregeln für das Arbeiten mit Mengen

Vereinigungs-Menge, Schnitt-Menge und die Differenz zweier Mengen genügen Gesetzmäßigkeiten, die in den folgenden Gleichungen zusammengefaßt sind. Bevor wir diese Gesetze angeben, vereinbaren wir noch, die leere Menge $\{\}$ mit \emptyset zu bezeichnen. Damit haben wir die folgenden Gesetze:

- | | |
|--------------------------------------------------------------------|-----------------------------------------------------------------|
| 1. $M \cup \emptyset = M$ | $M \cap \emptyset = \emptyset$ |
| 2. $M \cup M = M$ | $M \cap M = M$ |
| 3. $M \cup N = N \cup M$ | $M \cap N = N \cap M$ |
| 4. $(K \cup M) \cup N = K \cup (M \cup N)$ | $(K \cap M) \cap N = K \cap (M \cap N)$ |
| 5. $(K \cup M) \cap N = (K \cap N) \cup (M \cap N)$ | $(K \cap M) \cup N = (K \cup N) \cap (M \cup N)$ |
| 6. $M \setminus \emptyset = M$ | $M \setminus M = \emptyset$ |
| 7. $K \setminus (M \cup N) = (K \setminus M) \cap (K \setminus N)$ | $K \setminus (M \cap N) = (K \setminus M) \cup (K \setminus N)$ |
| 8. $(K \cup M) \setminus N = (K \setminus N) \cup (M \setminus N)$ | $(K \cap M) \setminus N = (K \setminus N) \cap (M \setminus N)$ |
| 9. $K \setminus (M \setminus N) = (K \setminus M) \cup (K \cap N)$ | $(K \setminus M) \setminus N = K \setminus (M \cup N)$ |
| 10. $M \cup (N \setminus M) = M \cup N$ | $M \cap (N \setminus M) = \emptyset$ |
| 11. $M \cup (M \cap N) = M$ | $M \cap (M \cup N) = M$ |

Wir beweisen exemplarisch die Gleichung $K \setminus (M \cup N) = (K \setminus M) \cap (K \setminus N)$. Um die Gleichheit zweier Mengen zu zeigen ist nachzuweisen, dass beide Mengen die selben Elemente enthalten. Wir haben die folgende Kette von Äquivalenzen:

$$\begin{aligned}
 & x \in K \setminus (M \cup N) \\
 \Leftrightarrow & x \in K \wedge \neg x \in M \cup N \\
 \Leftrightarrow & x \in K \wedge \neg (x \in M \vee x \in N) \\
 \Leftrightarrow & x \in K \wedge (\neg x \in M) \wedge (\neg x \in N) \\
 \Leftrightarrow & (x \in K \wedge \neg x \in M) \wedge (x \in K \wedge \neg x \in N) \\
 \Leftrightarrow & (x \in K \setminus M) \wedge (x \in K \setminus N) \\
 \Leftrightarrow & x \in (K \setminus M) \cap (K \setminus N).
 \end{aligned}$$

Die übrigen Gleichungen können nach dem selben Schema hergeleitet werden.

Zur Vereinfachung der Darstellung von Beweisen vereinbaren wir die folgende Schreibweise: Ist M eine Menge und x ein Objekt, so schreiben wir $x \notin M$ für die Formel $\neg x \in M$, formal:

$$x \notin M \stackrel{\text{def}}{\iff} \neg x \in M.$$

Eine analoge Notation verwenden wir auch für das Gleichheitszeichen:

$$x \neq y \stackrel{\text{def}}{\iff} \neg (x = y).$$

2.3 Binäre Relationen

Relationen treten in der Informatik an vielen Stellen auf. Die wichtigste Anwendung findet sich in der Theorie der relationalen Datenbanken. Wir betrachten im Folgenden den Spezialfall der *binären Relationen* und beleuchten das Verhältnis von binären Relationen und Funktionen. Wir werden sehen, dass wir Funktionen als spezielle binäre Relationen auffassen können. Damit stellt der Begriff der binären Relationen eine Verallgemeinerung des Funktions-Begriffs dar.

Zum Abschluß des Kapitels führen wir *transitive Relationen* und *Äquivalenz-Relationen* ein. Einerseits sind dies grundlegende Konzepte, die jeder Informatiker kennen muss, andererseits werden wir schon nächsten Kapitel mit Hilfe transitiver Relationen interessante Suchprobleme lösen können.

2.3.1 Binäre Relationen und Funktionen

Ist eine Menge R als Teilmenge des kartesischen Produkts zweier Mengen M und N gegeben, gilt also

$$R \subseteq M \times N,$$

so bezeichnen wir R auch als *binäre Relation*. In diesem Fall definieren wir den *Definitions-Bereich* von R als

$$\text{dom}(R) := \{x \mid \exists y \in N: \langle x, y \rangle \in R\}.$$

Entsprechend wird der *Werte-Bereich* von R als

$$\text{rng}(R) := \{y \mid \exists x \in M: \langle x, y \rangle \in R\}$$

definiert.

Beispiel: Es sei $R = \{\langle 1, 1 \rangle, \langle 2, 4 \rangle, \langle 3, 9 \rangle\}$. Dann gilt

$$\text{dom}(R) = \{1, 2, 3\} \quad \text{und} \quad \text{rng}(R) = \{1, 4, 9\}. \quad \square$$

Das nächste, stark vereinfachte Beispiel gibt einen Vorgeschmack von der Bedeutung binärer Relationen in der Theorie der *relationalen Datenbanken*.

Beispiel: Ein Autoverkäufer speichert in seiner Datenbank, welcher Kunde welches Auto gekauft hat. Nehmen wir an, dass die Mengen *Auto* und *Kunde* wie folgt gegeben sind:

$$\text{Kunde} = \{\text{Bauer, Maier, Schmidt}\} \quad \text{und} \quad \text{Auto} = \{\text{Polo, Fox, Golf}\}.$$

Dann könnte die binäre Relation

$$\text{Verkauf} \subseteq \text{Kunde} \times \text{Auto}$$

beispielsweise durch die folgende Menge gegeben sein:

$$\text{Verkauf} = \{\langle \text{Bauer, Golf} \rangle, \langle \text{Bauer, Fox} \rangle, \langle \text{Schmidt, Polo} \rangle\}.$$

Die Relation *Verkauf* drückt aus, dass der Kunde Bauer einen Golf und einen Fox erworben hat, der Kunde Schmidt hat einen Polo gekauft und Maier hat bisher noch kein Auto erworben. In der Theorie der Datenbanken werden Relationen üblicherweise in Form von Tabellen dargestellt. Die oben angegebene Relation hätte dann die folgende Form:

<i>Kunde</i>	<i>Auto</i>
Bauer	Golf
Bauer	Fox
Schmidt	Polo

Die oberste Zeile, in der wir die Spalten-Überschriften *Kunde* und *Auto* angeben, gehört selbst nicht zu der Relation, sondern wird als *Relationen-Schema* bezeichnet und die Relation zusammen mit ihrem Relationen-Schema nennen wir *Tabelle*.

Links- und Rechts-Eindeutige Relationen Wir nennen eine Relation $R \subseteq M \times N$ *rechts-eindeutig*, wenn folgendes gilt:

$$\forall x \in M: \forall y_1, y_2 \in N: (\langle x, y_1 \rangle \in R \wedge \langle x, y_2 \rangle \in R \rightarrow y_1 = y_2).$$

Bei einer rechts-eindeutigen Relation $R \subseteq M \times N$ gibt es also zu jedem $x \in M$ höchstens ein $y \in N$ so, dass $\langle x, y \rangle \in R$ gilt. Entsprechend nennen wir eine Relation $R \subseteq M \times N$ *links-eindeutig*, wenn gilt:

$$\forall y \in N: \forall x_1, x_2 \in M: (\langle x_1, y \rangle \in R \wedge \langle x_2, y \rangle \in R \rightarrow x_1 = x_2).$$

Bei einer links-eindeutigen Relation $R \subseteq M \times N$ gibt es also zu jedem $y \in N$ höchstens ein $x \in M$ so, dass $\langle x, y \rangle \in R$ gilt.

Beispiele: Es sei $M = \{1, 2, 3\}$ und $N = \{4, 5, 6\}$.

1. Die Relation R_1 sei definiert durch

$$R_1 = \{\langle 1, 4 \rangle, \langle 1, 6 \rangle\}.$$

Diese Relation ist nicht rechts-eindeutig, denn $4 \neq 6$. Die Relation ist links-eindeutig, denn die rechten Seiten aller in R_1 auftretenden Tupel sind verschieden.

2. Die Relation R_2 sei definiert durch

$$R_2 = \{\langle 1, 4 \rangle, \langle 2, 6 \rangle\}.$$

Diese Relation ist rechts-eindeutig, denn die linken Seiten aller in R_2 auftretenden Tupel sind verschieden. Sie ist auch links-eindeutig, denn die rechten Seiten aller in R_2 auftretenden Tupel sind verschieden.

3. Die Relation R_3 sei definiert durch

$$R_3 = \{\langle 1, 4 \rangle, \langle 2, 6 \rangle, \langle 3, 6 \rangle\}.$$

Diese Relation ist rechts-eindeutig, denn die linken Seiten aller in R_3 auftretenden Tupel sind verschieden. Sie ist nicht links-eindeutig, denn es gilt $\langle 2, 6 \rangle \in R$ und $\langle 3, 6 \rangle \in R$, aber $2 \neq 3$. \diamond

Totale Relationen Eine binäre Relation $R \subseteq M \times N$ heißt *links-total auf M* , wenn

$$\forall x \in M: \exists y \in N: \langle x, y \rangle \in R$$

gilt. Dann gibt es für alle x aus der Menge M ein y aus der Menge N , so dass $\langle x, y \rangle$ in der Menge R liegt. Die Relation R_3 aus dem obigen Beispiel ist links-total, denn jedem Element aus M wird durch R_3 ein Element aus N zugeordnet.

Analog nennen wir eine binäre Relation $R \subseteq M \times N$ *rechts-total auf N* , wenn

$$\forall y \in N: \exists x \in M: \langle x, y \rangle \in R$$

gilt. Dann gibt es für alle y aus der Menge N ein x aus der Menge M , so dass $\langle x, y \rangle$ in der Menge R liegt. Die Relation R_3 aus dem obigen Beispiel ist nicht rechts-total, denn dem Element 5 aus N wird durch R_3 kein Element aus M zugeordnet, denn für alle $\langle x, y \rangle \in R_3$ gilt $y \neq 5$.

Funktionale Relationen Eine Relation $R \subseteq M \times N$, die sowohl links-total auf M als auch rechts-eindeutig ist, nennen wir eine *funktionale* Relation auf M . Ist $R \subseteq M \times N$ eine funktionale Relation, so können wir eine Funktion $f_R: M \rightarrow N$ wie folgt definieren:

$$f_R(x) := y \stackrel{\text{def}}{\iff} \langle x, y \rangle \in R.$$

Diese Definition funktioniert, denn aus der Links-Totalität von R folgt, dass es für jedes $x \in M$ auch ein $y \in N$ gibt, so dass $\langle x, y \rangle \in R$ ist. Aus der Rechts-Eindeutigkeit von R folgt dann, dass dieses y eindeutig bestimmt ist. Ist umgekehrt eine Funktion $f: M \rightarrow N$ gegeben, so können wir dieser Funktion eine Relation $\text{graph}(f) \subseteq M \times N$ zuordnen, indem wir definieren:

$$\text{graph}(f) := \{\langle x, f(x) \rangle \mid x \in M\}.$$

Die so definierte Relation $\text{graph}(f)$ ist links-total, denn die Funktion f berechnet ja für jedes $x \in M$ ein Ergebnis und die Relation ist rechts-eindeutig, denn die Funktion berechnet ja für ein Argument nicht mehrere verschiedene Ergebnisse.

Aufgrund der gerade diskutierten Korrespondenz zwischen Funktionen und Relationen werden wir daher im Folgenden alle Funktionen als spezielle binäre Relationen auffassen. Für die Menge aller Funktionen von M nach N schreiben wir auch N^M , genauer definieren wir

$$N^M := \{R \subseteq M \times N \mid R \text{ funktional}\}.$$

Diese Schreibweise erklärt sich wie folgt: Sind M und N endliche Mengen mit m bzw. n Elementen, so gibt es genau n^m verschiedene Funktionen von M nach N , es gilt also

$$\text{card}(N^M) = \text{card}(N)^{\text{card}(M)}.$$

Wir werden daher funktionale Relationen und die entsprechenden Funktionen identifizieren. Damit ist dann für eine funktionale Relation $R \subseteq M \times N$ und ein $x \in M$ auch die Schreibweise $R(x)$ zulässig: Mit $R(x)$ bezeichnen wir das eindeutig bestimmte $y \in N$, für das $\langle x, y \rangle \in R$ gilt.

Beispiele:

1. Wir setzen $M = \{1, 2, 3\}$, $N = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ und definieren

$$R := \{\langle 1, 1 \rangle, \langle 2, 4 \rangle, \langle 3, 9 \rangle\}.$$

Dann ist R eine funktionale Relation auf M . Diese Relation berechnet gerade die Quadrat-Zahlen auf der Menge M .

2. Diesmal setzen wir $M = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ und $N = \{1, 2, 3\}$ und definieren

$$R := \{\langle 1, 1 \rangle, \langle 4, 2 \rangle, \langle 9, 3 \rangle\}.$$

Dann ist R keine funktionale Relation auf M , denn R ist nicht links-total auf M . Beispielsweise wird das Element 2 von der Relation R auf kein Element aus N abgebildet.

3. Wir setzen nun $M = \{1, 2, 3\}$, $N = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ und definieren

$$R := \{\langle 1, 1 \rangle, \langle 2, 3 \rangle, \langle 2, 4 \rangle, \langle 3, 9 \rangle\}$$

Dann ist R keine funktionale Relation auf M , denn R ist nicht rechts-eindeutig auf M . Beispielsweise wird das Element 2 von der Relation R sowohl auf 3 als auch auf 4 abgebildet. \diamond

Ist $R \subseteq M \times N$ eine binäre Relation und ist weiter $X \subseteq M$, so definieren wir das *Bild von X unter R* als

$$R(X) := \{y \mid \exists x \in X: \langle x, y \rangle \in R\}.$$

Inverse Relation Zu einer Relation $R \subseteq M \times N$ definieren wir die *inverse* Relation $R^{-1} \subseteq N \times M$ wie folgt:

$$R^{-1} := \{\langle y, x \rangle \mid \langle x, y \rangle \in R\}.$$

Aus dieser Definition folgt sofort, dass R^{-1} rechts-eindeutig ist genau dann, wenn R links-eindeutig ist. Außerdem ist R^{-1} links-total genau dann, wenn R rechts-total ist. Ist eine Relation sowohl links-eindeutig als auch rechts-eindeutig und außerdem sowohl links-total als auch rechts-total, so nennen wir sie auch *bijektiv*. In diesem Fall läßt sich neben der Funktion f_R auch eine Funktion $f_{R^{-1}}$ definieren. Die Definition der letzten Funktion lautet ausgeschrieben:

$$f_{R^{-1}}(y) := x \stackrel{\text{def}}{\iff} \langle y, x \rangle \in R^{-1} \iff \langle x, y \rangle \in R.$$

Diese Funktion ist dann aber genau die Umkehr-Funktion von f_R , es gilt

$$\forall y \in N: f_R(f_{R^{-1}}(y)) = y \quad \text{und} \quad \forall x \in M: f_{R^{-1}}(f_R(x)) = x.$$

Dieser Umstand rechtfertigt im nachhinein die Schreibweise R^{-1} .

Komposition von Relationen Ähnlich wie wir Funktionen verknüpfen können, können auch Relationen verknüpft werden. Wir betrachten zunächst Mengen L , M und N . Sind dort zwei Relationen $R \subseteq L \times M$ und $Q \subseteq M \times N$ definiert, so ist das *relationale Produkt* $R \circ Q$ wie folgt definiert:

$$R \circ Q := \{\langle x, z \rangle \mid \exists y \in M: (\langle x, y \rangle \in R \wedge \langle y, z \rangle \in Q)\}$$

Das relationale Produkt von Q und R wird gelegentlich auch als die *Komposition* von Q und R bezeichnet.

Beispiel: Es sei $L = \{1, 2, 3\}$, $M = \{4, 5, 6\}$ und $N = \{7, 8, 9\}$. Weiter seien die Relationen Q und R wie folgt gegeben:

$$R = \{\langle 1, 4 \rangle, \langle 1, 6 \rangle, \langle 3, 5 \rangle\} \quad \text{und} \quad Q = \{\langle 4, 7 \rangle, \langle 6, 8 \rangle, \langle 6, 9 \rangle\}.$$

Dann gilt

$$R \circ Q = \{\langle 1, 7 \rangle, \langle 1, 8 \rangle, \langle 1, 9 \rangle\}.$$

◇

Ist $R \subseteq L \times M$ eine funktionale Relation auf L und ist $Q \subseteq M \times N$ eine funktionale Relation auf M , so ist auch $R \circ Q$ eine funktionale Relation auf L und die Funktion $f_{R \circ Q}$ kann wie folgt aus den Funktionen f_R und f_Q berechnet werden:

$$f_{R \circ Q}(x) = f_Q(f_R(x)).$$

Bemerkung: In einigen Lehrbüchern wird das relationale Produkt, das wir als $R \circ Q$ definiert haben, mit $Q \circ R$ bezeichnet. Damit lautet die Definition von $R \circ Q$ dann wie folgt: Ist $R \subseteq M \times N$ und $Q \subseteq L \times M$, dann ist

$$R \circ Q := \{\langle x, z \rangle \mid \exists y \in M: (\langle x, y \rangle \in Q \wedge \langle y, z \rangle \in R)\}.$$

Diese Definition hat den folgenden Vorteil: Falls R und Q funktionale Relationen sind und wenn dann weiter f und g die diesen Relationen zugeordneten Funktionen sind, wenn also

$$Q = \text{graph}(f) \quad \text{und} \quad R = \text{graph}(g)$$

gilt, dann haben wir für die Komposition der Funktionen f und g , die durch $(g \circ f)(x) = g(f(x))$ definiert ist, die Gleichung

$$\text{graph}(g \circ f) = R \circ Q = \text{graph}(g) \circ \text{graph}(f).$$

Die von uns verwendete Definition hat den Vorteil, dass später die Berechnung des transitiven Abschlusses einer Relation intuitiver wird. ◇

Beispiel: Das nächste Beispiel zeigt die Verwendung des relationalen Produkts im Kontext einer Datenbank. Wir nehmen an, dass die Datenbank eines Autohändlers unter anderem die folgenden beiden Tabellen enthält.

Kauf:

Kunde	Auto
Bauer	Golf
Bauer	Fox
Schmidt	Polo

Preis:

Auto	Betrag
Golf	20 000
Fox	10 000
Polo	13 000

Dann ist das relationale Produkt der in den Tabellen *Kauf* und *Preis* dargestellten Relationen durch die in der folgenden Tabelle dargestellten Relation gegeben:

Kunde	Betrag
Bauer	20 000
Bauer	10 000
Schmidt	13 000

Diese Relation könnte dann zur Rechnungsstellung weiter verwendet werden. ◇

Bemerkung: In der Theorie der Datenbanken wird der Begriff der Komposition zweier Relationen zu dem Begriff des *Joins* verallgemeinert. ◇

Eigenschaften des relationalen Produkts Die Komposition von Relationen ist *assoziativ*: Sind

$$R \subseteq K \times L, \quad Q \subseteq L \times M \quad \text{und} \quad P \subseteq M \times N$$

binäre Relationen, so gilt

$$(R \circ Q) \circ P = R \circ (Q \circ P).$$

Beweis: Wir zeigen

$$\langle x, u \rangle \in (R \circ Q) \circ P \leftrightarrow \langle x, u \rangle \in R \circ (Q \circ P) \quad (2.2)$$

Wir formen zunächst die linke Seite $\langle x, u \rangle \in (R \circ Q) \circ P$ der Äquivalenz 2.2 um. Es gilt

$$\begin{aligned} & \langle x, u \rangle \in (R \circ Q) \circ P \\ \Leftrightarrow & \exists z : (\langle x, z \rangle \in R \circ Q \wedge \langle z, u \rangle \in P) && \text{nach Def. von } (R \circ Q) \circ P \\ \Leftrightarrow & \exists z : ((\exists y : \langle x, y \rangle \in R \wedge \langle y, z \rangle \in Q) \wedge \langle z, u \rangle \in P) && \text{nach Def. von } R \circ Q \end{aligned}$$

Da die Variable y in der Formel $\langle z, u \rangle \in P$ nicht auftritt, können wir den Existenz-Quantor über y auch herausziehen:

$$\exists z : \exists y : (\langle x, y \rangle \in R \wedge \langle y, z \rangle \in Q \wedge \langle z, u \rangle \in P) \quad (2.3)$$

Wir formen nun die rechte Seite der Äquivalenz 2.2 um:

$$\begin{aligned} & \langle x, u \rangle \in R \circ (Q \circ P) \\ \Leftrightarrow & \exists y : (\langle x, y \rangle \in R \wedge \langle y, u \rangle \in Q \circ P) && \text{nach Def. von } R \circ (Q \circ P) \\ \Leftrightarrow & \exists y : (\langle x, y \rangle \in R \wedge \exists z : (\langle y, z \rangle \in Q \wedge \langle z, u \rangle \in P)) && \text{nach Def. von } Q \circ P \end{aligned}$$

Da die Variable z in der Formel $\langle x, y \rangle \in R$ nicht vorkommt, können wir den Existenz-Quantor über z auch vorziehen und haben dann

$$\exists y : \exists z : (\langle x, y \rangle \in R \wedge \langle y, z \rangle \in Q \wedge \langle z, u \rangle \in P) \quad (2.4)$$

Die Formeln (2.3) und (2.4) unterscheiden sich nur durch die Reihenfolge der Existenz-Quantoren und sind daher äquivalent. Damit ist die Äquivalenz (2.2) nachgewiesen und der Beweis der Assoziativität des Kompositions-Operators ist erbracht. \square

Eine weitere wichtige Eigenschaft des relationalen Produkts ist die folgende: Sind zwei Relationen $R \subseteq L \times M$ und $Q \subseteq M \times N$ gegeben, so gilt

$$(R \circ Q)^{-1} = Q^{-1} \circ R^{-1}.$$

Beachten Sie, dass sich die Reihenfolgen von Q und R hier vertauschen. Zum Beweis ist zu zeigen, dass für alle Paare $\langle z, x \rangle \in N \times L$ die Äquivalenz

$$\langle z, x \rangle \in (Q \circ R)^{-1} \Leftrightarrow \langle z, x \rangle \in R^{-1} \circ Q^{-1}$$

gilt. Den Nachweis erbringen wir durch die folgende Kette von Äquivalenz-Umformungen:

$$\begin{aligned} & \langle z, x \rangle \in (R \circ Q)^{-1} \\ \Leftrightarrow & \langle x, z \rangle \in R \circ Q \\ \Leftrightarrow & \exists y \in M : (\langle x, y \rangle \in R \wedge \langle y, z \rangle \in Q) \\ \Leftrightarrow & \exists y \in M : (\langle y, z \rangle \in Q \wedge \langle x, y \rangle \in R) \\ \Leftrightarrow & \exists y \in M : (\langle z, y \rangle \in Q^{-1} \wedge \langle y, x \rangle \in R^{-1}) \\ \Leftrightarrow & \langle z, x \rangle \in Q^{-1} \circ R^{-1} \quad \square \end{aligned}$$

Wir bemerken noch, dass das folgende Distributivgesetz gilt: Sind R_1 und R_2 Relationen auf $L \times M$ und ist Q eine Relation auf $M \times N$, so gilt

$$(R_1 \cup R_2) \circ Q = (R_1 \circ Q) \cup (R_2 \circ Q).$$

Analog gilt ebenfalls

$$R \circ (Q_1 \cup Q_2) = (R \circ Q_1) \cup (R \circ Q_2),$$

falls R eine Relation auf $L \times M$ und Q_1 und Q_2 Relationen auf $M \times N$ sind. Um Gleichungen der obigen Art kürzer schreiben zu können vereinbaren wir, dass der Kompositions-Operator \circ stärker bindet als \cup und \cap . Wir beweisen nun das erste Distributivgesetz, indem wir

$$\langle x, z \rangle \in (R_1 \cup R_2) \circ Q \Leftrightarrow \langle x, z \rangle \in R_1 \circ Q \cup R_2 \circ Q \quad (2.5)$$

zeigen. Wir formen zunächst den Ausdruck $\langle x, z \rangle \in (R_1 \cup R_2) \circ Q$ um:

$$\begin{aligned} & \langle x, z \rangle \in (R_1 \cup R_2) \circ Q \\ \leftrightarrow & \exists y : (\langle x, y \rangle \in R_1 \cup R_2 \wedge \langle y, z \rangle \in Q) && \text{nach Def. von } (R_1 \cup R_2) \circ Q \\ \leftrightarrow & \exists y : ((\langle x, y \rangle \in R_1 \vee \langle x, y \rangle \in R_2) \wedge \langle y, z \rangle \in Q) && \text{nach Def. von } R_1 \cup R_2 \end{aligned}$$

Diese Formel stellen wir mit Hilfe des Distributiv-Gesetzes der Aussagen-Logik um. In der Aussagenlogik werden wir später zeigen dass für beliebige Formeln F_1 , F_2 und G die Äquivalenz

$$(F_1 \vee F_2) \wedge G \leftrightarrow (F_1 \wedge G) \vee (F_2 \wedge G)$$

gilt. Die Anwendung dieses Gesetzes liefert:

$$\begin{aligned} & \exists y : \left(\underbrace{(\langle x, y \rangle \in R_1)}_{F_1} \vee \underbrace{(\langle x, y \rangle \in R_2)}_{F_2} \wedge \underbrace{\langle y, z \rangle \in Q}_G \right) \\ \leftrightarrow & \exists y : \left(\underbrace{(\langle x, y \rangle \in R_1 \wedge \langle y, z \rangle \in Q)}_{F_1 \wedge G} \vee \underbrace{(\langle x, y \rangle \in R_2 \wedge \langle y, z \rangle \in Q)}_{F_2 \wedge G} \right) \end{aligned} \quad (2.6)$$

Wir formen nun den Ausdruck $\langle x, z \rangle \in R_1 \circ Q \cup R_2 \circ Q$ um:

$$\begin{aligned} & \langle x, z \rangle \in R_1 \circ Q \cup R_2 \circ Q \\ \leftrightarrow & \langle x, z \rangle \in R_1 \circ Q \vee \langle x, z \rangle \in R_2 \circ Q && \text{nach Def. von } \cup \\ \leftrightarrow & (\exists y : (\langle x, y \rangle \in R_1 \wedge \langle y, z \rangle \in Q)) \vee (\exists y : (\langle x, y \rangle \in R_2 \wedge \langle y, z \rangle \in Q)) \\ & \text{nach Def. von } R_1 \circ Q \text{ und } R_2 \circ Q \end{aligned}$$

Diese letzte Formel stellen wir mit Hilfe eines Distributiv-Gesetzes für die Prädikaten-Logik um. In der Prädikaten-Logik werden wir später sehen, dass für beliebige Formeln F_1 und F_2 die Äquivalenz

$$\exists y : (F_1 \vee F_2) \leftrightarrow (\exists y : F_1) \vee (\exists y : F_2)$$

gültig ist. Damit folgt dann

$$\begin{aligned} & \exists y : \left(\underbrace{(\langle x, y \rangle \in R_1 \wedge \langle y, z \rangle \in Q)}_{F_1} \vee \underbrace{(\langle x, y \rangle \in R_2 \wedge \langle y, z \rangle \in Q)}_{F_2} \right) \\ \leftrightarrow & \exists y : \left(\underbrace{(\langle x, y \rangle \in R_1 \wedge \langle y, z \rangle \in Q)}_{F_1} \vee \underbrace{(\langle x, y \rangle \in R_2 \wedge \langle y, z \rangle \in Q)}_{F_2} \right) \end{aligned} \quad (2.7)$$

Da die Formeln 2.6 und 2.7 identisch sind, ist der Beweis des Distributiv-Gesetzes $(R_1 \cup R_2) \circ Q = R_1 \circ Q \cup R_2 \circ Q$ erbracht. \square

Interessant ist noch zu bemerken, dass für den Schnitt von Relationen und dem Kompositions-Operator kein Distributivgesetz gilt, die Gleichung

$$(R_1 \cap R_2) \circ Q = R_1 \circ Q \cap R_2 \circ Q$$

ist im Allgemeinen falsch. Um diese Behauptung zu belegen, benötigen wir ein Gegenbeispiel. Dazu definieren wir die Relationen R_1 , R_2 und Q wie folgt:

$$R_1 := \{\langle 1, 2 \rangle\}, \quad R_2 := \{\langle 1, 3 \rangle\} \quad \text{und} \quad Q = \{\langle 2, 4 \rangle, \langle 3, 4 \rangle\}.$$

Dann gilt

$$R_1 \circ Q = \{\langle 1, 4 \rangle\}, \quad R_2 \circ Q = \{\langle 1, 4 \rangle\}, \quad \text{also } R_1 \circ Q \cap R_2 \circ Q = \{\langle 1, 4 \rangle\},$$

aber andererseits haben wir

$$(R_1 \cap R_2) \circ Q = \emptyset \circ Q = \emptyset \neq \{\langle 1, 4 \rangle\} = R_1 \circ Q \cap R_2 \circ Q.$$

Identische Relation Ist M eine Menge, so definieren wir die *identische Relation* $\text{id}_M \subseteq M \times M$ wie folgt:

$$\text{id}_M := \{\langle x, x \rangle \mid x \in M\}.$$

Beispiel: Es sei $M = \{1, 2, 3\}$. Dann gilt

$$\text{id}_M := \{\langle 1, 1 \rangle, \langle 2, 2 \rangle, \langle 3, 3 \rangle\}.$$

Aus der Definition folgt sofort

$$\text{id}_M^{-1} = \text{id}_M.$$

Sei weiterhin $R \subseteq M \times N$ und eine binäre Relation, so gilt

$$R \circ \text{id}_N = R \quad \text{und} \quad \text{id}_M \circ R = R.$$

Wir weisen die zweite Gleichung nach. Nach Definition des relationalen Produkts gilt

$$\text{id}_M \circ R = \{\langle x, z \rangle \mid \exists y : \langle x, y \rangle \in \text{id}_M \wedge \langle y, z \rangle \in R\}.$$

Nun ist $\langle x, y \rangle \in \text{id}_M$ genau dann, wenn $x = y$ ist, also gilt

$$\text{id}_M \circ R = \{\langle x, z \rangle \mid \exists y : x = y \wedge \langle y, z \rangle \in R\}.$$

Es gilt die folgende Äquivalenz

$$(\exists y : x = y \wedge \langle y, z \rangle \in R) \leftrightarrow \langle x, z \rangle \in R.$$

Diese Äquivalenz ist leicht einzusehen: Falls $\exists y : x = y \wedge \langle y, z \rangle \in R$ gilt, so muß dass y dessen Existenz gefordert wird, den Wert x haben und dann gilt natürlich auch $\langle x, z \rangle \in R$. Gilt andererseits $\langle x, z \rangle \in R$, so definieren wir $y := x$. Für das so definierte y gilt offensichtlich $x = y \wedge \langle y, z \rangle \in R$. Unter Verwendung der oberen Äquivalenz haben wir

$$\text{id}_M \circ R = \{\langle x, z \rangle \mid \langle x, z \rangle \in R\}.$$

Wegen $R \subseteq M \times N$ besteht R nur aus geordneten Paaren und daher gilt

$$R = \{\langle x, z \rangle \mid \langle x, z \rangle \in R\}.$$

Damit ist $\text{id}_M \circ R = R$ gezeigt. □

Aufgabe 1: Es sei $R \subseteq M \times N$. Unter welchen Bedingungen gilt

$$R \circ R^{-1} = \text{id}_M?$$

Unter welchen Bedingungen gilt

$$R^{-1} \circ R = \text{id}_M? \quad \diamond$$

2.3.2 Binäre Relationen auf einer Menge

Wir betrachten im folgenden den Spezialfall von Relationen $R \subseteq M \times N$, für den $M = N$ gilt. Wir definieren: Eine Relation $R \subseteq M \times M$ heißt eine Relation *auf* der Menge M . Im Rest dieses Abschnittes betrachten wir nur noch solche Relationen. Statt $M \times M$ schreiben wir auch kürzer M^2 .

Ist R eine Relation auf M und sind $x, y \in M$, so schreiben wir statt $\langle x, y \rangle \in R$ gelegentlich auch $x R y$. Beispielsweise läßt sich die Relation \leq auf \mathbb{N} wie folgt definieren:

$$\leq := \{\langle x, y \rangle \in \mathbb{N} \times \mathbb{N} \mid \exists z \in \mathbb{N} : x + z = y\}.$$

Statt $\langle x, y \rangle \in \leq$ hat sich die Schreibweise $x \leq y$ eingebürgert.

Definition 1 Eine Relation $R \subseteq M \times M$ ist reflexiv auf der Menge M falls gilt:

$$\forall x \in M : \langle x, x \rangle \in R.$$

Satz 2 Eine Relation $R \subseteq M \times M$ ist genau dann reflexiv, wenn $\text{id}_M \subseteq R$ gilt.

Beweis: Es gilt

$$\text{id}_M \subseteq R$$

$$\text{g.d.w. } \{\langle x, x \rangle \mid x \in M\} \subseteq R$$

$$\text{g.d.w. } \forall x \in M : \langle x, x \rangle \in R$$

$$\text{g.d.w. } R \text{ reflexiv}$$

□

Definition 3 Eine Relation $R \subseteq M \times M$ ist symmetrisch falls gilt:

$$\forall x, y \in M : (\langle x, y \rangle \in R \rightarrow \langle y, x \rangle \in R).$$

Satz 4 Eine Relation $R \subseteq M \times M$ ist genau dann symmetrisch, wenn $R^{-1} \subseteq R$ gilt.

Beweis: Die Äquivalenz der beiden Bedingungen wird offensichtlich, wenn wir die Inklusions-Bedingung $R^{-1} \subseteq R$ expandieren, indem wir die Gleichungen

$$R^{-1} = \{\langle y, x \rangle \mid \langle x, y \rangle \in R\} \quad \text{und} \quad R = \{\langle x, y \rangle \mid \langle x, y \rangle \in R\}$$

berücksichtigen, denn dann hat die Inklusions-Bedingung die Form

$$\{\langle y, x \rangle \mid \langle x, y \rangle \in R\} \subseteq \{\langle x, y \rangle \mid \langle x, y \rangle \in R\}.$$

Nach der Definition der Teilmengen-Beziehung ist diese Bedingung gleichwertig zu der Formel

$$\forall x, y \in M : (\langle x, y \rangle \in R \rightarrow \langle y, x \rangle \in R).$$

□

Definition 5 Eine Relation $R \subseteq M \times M$ ist anti-symmetrisch falls gilt:

$$\forall x, y \in M : (\langle x, y \rangle \in R \wedge \langle y, x \rangle \in R \rightarrow x = y).$$

Satz 6 Eine Relation $R \subseteq M \times M$ ist genau dann anti-symmetrisch, wenn $R \cap R^{-1} \subseteq \text{id}_M$ gilt.

Beweis: Wir nehmen zunächst an, dass R anti-symmetrisch ist und folglich

$$\forall x, y \in M : \langle x, y \rangle \in R \wedge \langle y, x \rangle \in R \rightarrow x = y$$

gilt und zeigen, dass aus dieser Voraussetzung die Inklusions-Beziehung $R \cap R^{-1} \subseteq \text{id}_M$ folgt. Sei also $\langle x, y \rangle \in R \cap R^{-1}$. Dann gilt einerseits $\langle x, y \rangle \in R$ und andererseits folgt aus $\langle x, y \rangle \in R^{-1}$ dass auch $\langle y, x \rangle \in R$ ist. Dann folgt aber aus der Voraussetzung sofort $x = y$ und das impliziert $\langle x, y \rangle \in \text{id}_M$, womit $R \cap R^{-1} \subseteq \text{id}_M$ gezeigt ist.

Wir nehmen nun an, dass $R \cap R^{-1} \subseteq \text{id}_M$ gilt und zeigen, dass daraus die Gültigkeit von

$$\forall x, y \in M : \langle x, y \rangle \in R \wedge \langle y, x \rangle \in R \rightarrow x = y$$

folgt. Seien also $x, y \in M$ und es gelte $\langle x, y \rangle \in R$ und $\langle y, x \rangle \in R$. Wir müssen zeigen, dass daraus $x = y$ folgt. Aus $\langle y, x \rangle \in R$ folgt $\langle x, y \rangle \in R^{-1}$. Also gilt $\langle x, y \rangle \in R \cap R^{-1}$. Aus der Inklusions-Beziehung $R \cap R^{-1} \subseteq \text{id}_M$ folgt dann $\langle x, y \rangle \in \text{id}_M$ und daraus folgt sofort $x = y$. □

Definition 7 Eine Relation $R \subseteq M \times M$ ist transitiv falls gilt:

$$\forall x, y, z \in M : \langle x, y \rangle \in R \wedge \langle y, z \rangle \in R \rightarrow \langle x, z \rangle \in R.$$

Satz 8 Eine Relation $R \subseteq M \times M$ ist genau dann transitiv, wenn $R \circ R \subseteq R$ ist.

Beweis: Wir nehmen zunächst an, dass R transitiv ist und damit

$$\forall x, y, z \in M : \langle x, y \rangle \in R \wedge \langle y, z \rangle \in R \rightarrow \langle x, z \rangle \in R$$

gilt und zeigen, dass daraus $R \circ R \subseteq R$ folgt. Sei also $\langle x, z \rangle \in R \circ R$. Nach Definition des relationalen Produkts gibt es dann ein y , so dass $\langle x, y \rangle \in R$ und $\langle y, z \rangle \in R$ gilt. Nach Voraussetzung gilt jetzt $\langle x, z \rangle \in R$ und das war zu zeigen.

Wir nehmen nun an, dass die Inklusion $R \circ R \subseteq R$ gilt und zeigen, dass daraus

$$\forall x, y, z \in M : \langle x, y \rangle \in R \wedge \langle y, z \rangle \in R \rightarrow \langle x, z \rangle \in R$$

folgt. Seien also $x, y, z \in M$ mit $\langle x, y \rangle \in R$ und $\langle y, z \rangle \in R$ gegeben. Nach Definition des relationalen Produkts gilt dann $\langle x, z \rangle \in R \circ R$ und aus der Voraussetzung $R \circ R \subseteq R$ folgt $\langle x, z \rangle \in R$. \square

Beispiele: In den ersten beiden Beispielen sei $M = \{1, 2, 3\}$.

1. $R_1 = \{\langle 1, 1 \rangle, \langle 2, 2 \rangle, \langle 3, 3 \rangle\}$.

R_1 ist reflexiv, symmetrisch, anti-symmetrisch und transitiv.

2. $R_2 = \{\langle 1, 2 \rangle, \langle 2, 1 \rangle, \langle 3, 3 \rangle\}$.

R_2 ist nicht reflexiv, da $\langle 1, 1 \rangle \notin R_2$. R_2 ist symmetrisch. R_2 ist nicht anti-symmetrisch, denn aus $\langle 1, 2 \rangle \in R_2$ und $\langle 2, 1 \rangle \in R_2$ müßte $2 = 1$ folgen. Schließlich ist R_2 auch nicht transitiv, denn aus $\langle 1, 2 \rangle \in R_2$ und $\langle 2, 1 \rangle \in R_2$ müßte $\langle 1, 1 \rangle \in R_2$ folgen.

In den beiden folgenden Beispielen sei $M = \mathbb{N}$.

3. $R_3 := \{\langle n, m \rangle \in \mathbb{N}^2 \mid n \leq m\}$.

R_3 ist reflexiv, denn für alle natürlichen Zahlen $n \in \mathbb{N}$ gilt $n \leq n$. R_3 ist nicht symmetrisch, denn beispielsweise gilt $1 \leq 2$, aber es gilt nicht $2 \leq 1$. Allerdings ist R_3 anti-symmetrisch, denn wenn $n \leq m$ und $m \leq n$ gilt, so muß schon $m = n$ gelten. Schließlich ist R_3 auch transitiv, denn aus $k \leq m$ und $m \leq n$ folgt natürlich $k \leq n$.

4. $R_4 := \{\langle m, n \rangle \in \mathbb{N}^2 \mid \exists k \in \mathbb{N} : m \cdot k = n\}$

Für zwei positive Zahlen m und n gilt $\langle m, n \rangle \in R_4$ genau dann, wenn m ein Teiler von n ist. Damit ist klar, dass R_4 reflexiv ist, denn jede Zahl teilt sich selbst. Natürlich ist R_4 nicht symmetrisch, denn 1 ist ein Teiler von 2 aber nicht umgekehrt. Dafür ist R_4 aber anti-symmetrisch, denn wenn sowohl m ein Teiler von n ist und auch n ein Teiler von m , so muß $m = n$ gelten. Schließlich ist R_4 auch transitiv: Ist m ein Teiler von n und n ein Teiler von o , so ist natürlich m ebenfalls ein Teiler von o . \diamond

Ist R eine Relation auf M , die nicht transitiv ist, so können wir R zu einer transitiven Relation erweitern. Dazu definieren wir für alle $n \in \mathbb{N}$ die Potenzen R^n durch Induktion über n .

1. Induktions-Anfang: $n = 0$. Wir setzen

$$R^0 := \text{id}_M$$

2. Induktions-Schritt: $n \rightarrow n + 1$. Nach Induktions-Voraussetzung ist R^n bereits definiert. Daher können wir R^{n+1} definieren als

$$R^{n+1} = R \circ R^n.$$

Wir benötigen später das folgende Potenz-Gesetz: Für beliebige natürliche Zahlen $k, l \in \mathbb{N}$ gilt:

$$R^k \circ R^l = R^{k+l}.$$

Beweis: Wir führen den Beweis durch Induktion nach k .

I.A.: $k = 0$. Es gilt

$$R^0 \circ R^l = \text{id}_M \circ R^l = R^l = R^{0+l}.$$

I.S.: $k \mapsto k + 1$. Es gilt

$$\begin{aligned} R^{k+1} \circ R^l &= (R \circ R^k) \circ R^l && \text{nach Def. von } R^{k+1} \\ &= R \circ (R^k \circ R^l) && \text{aufgrund des Assoziativ-Gesetzes für } \circ \\ &= R \circ R^{k+l} && \text{nach Induktions-Voraussetzung} \\ &= R^{(k+l)+1} && \text{nach Def. von } R^{n+1} \\ &= R^{(k+1)+l}. && \square \end{aligned}$$

Wir definieren den *transitiven Abschluß* von R als die Menge

$$R^+ := \bigcup_{n=1}^{\infty} R^n.$$

Dabei ist für eine Folge $(A_n)_n$ von Mengen der Ausdruck $\bigcup_{i=1}^{\infty} A_n$ wie folgt definiert:

$$\bigcup_{i=1}^{\infty} A_n = A_1 \cup A_2 \cup A_3 \cup \dots$$

Satz 9 *Es sei M eine Menge und $R \subseteq M \times M$ eine binäre Relation auf M . Dann hat die oben definierte Relation R^+ die folgenden Eigenschaften:*

1. R^+ ist transitiv.
2. R^+ ist die bezüglich der Inklusions-Ordnung \subseteq kleinste Relation T auf M , die einerseits transitiv ist und andererseits die Relation R enthält. Anders ausgedrückt: Ist T eine transitive Relation auf M mit $R \subseteq T$, so muß $R^+ \subseteq T$ gelten.

Beweis:

1. Wir zeigen zunächst, dass R^+ transitiv ist. Dazu müssen wir die Gültigkeit der Formel

$$\forall x, y, z : \langle x, y \rangle \in R^+ \wedge \langle y, z \rangle \in R^+ \rightarrow \langle x, z \rangle \in R^+$$

nachweisen. Wir nehmen also an, dass $\langle x, y \rangle \in R^+$ und $\langle y, z \rangle \in R^+$ gilt und zeigen, dass aus dieser Voraussetzung auf $\langle x, z \rangle \in R^+$ geschlossen werden kann. Nach Definition von R^+ haben wir

$$\langle x, y \rangle \in \bigcup_{n=1}^{\infty} R^n \quad \text{und} \quad \langle y, z \rangle \in \bigcup_{n=1}^{\infty} R^n.$$

Nach der Definition der Menge $\bigcup_{n=1}^{\infty} R^n$ gibt es dann natürliche Zahlen $k, l \in \mathbb{N}$, so dass

$$\langle x, y \rangle \in R^k \quad \text{und} \quad \langle y, z \rangle \in R^l$$

gilt. Aus der Definition des relationalen Produktes folgt nun

$$\langle x, z \rangle \in R^k \circ R^l.$$

Aufgrund des Potenz-Gesetzes für das relationale Produkt gilt

$$R^k \circ R^l = R^{k+l}.$$

Also haben wir $\langle x, z \rangle \in R^{k+l}$ und daraus folgt sofort

$$\langle x, z \rangle \in \bigcup_{n=1}^{\infty} R^n.$$

Damit gilt $\langle x, z \rangle \in R^+$ und das war zu zeigen.

2. Um zu zeigen, dass R^+ die kleinste Relation ist, die einerseits transitiv ist und andererseits R enthält, nehmen wir an, dass T eine transitive Relation ist, für die $R \subseteq T$ gilt. Wir müssen dann zeigen, dass $R^+ \subseteq T$ gilt. Wir zeigen zunächst durch vollständige Induktion über $n \in \mathbb{N}$, dass für alle positiven natürlichen Zahlen $n \in \mathbb{N}$ die folgende Inklusion gilt:

$$R^n \subseteq T.$$

I.A.: $n = 1$. Dann ist $R^1 \subseteq T$ zu zeigen. Wegen $R^1 = R \circ \text{id}_M = R$ folgt dies aber unmittelbar aus der Voraussetzung $R \subseteq T$.

I.S.: $n \mapsto n + 1$. Nach Induktions-Voraussetzung wissen wir

$$R^n \subseteq T.$$

Wir multiplizieren diese Inklusion auf beiden Seiten relational mit R und haben dann

$$R^{n+1} = R \circ R^n \subseteq R \circ T.$$

Multiplizieren wir die Voraussetzung $R \subseteq T$ relational mit T , so finden wir

$$R \circ T \subseteq T \circ T.$$

Weil T transitiv ist, gilt

$$T \circ T \subseteq T.$$

Insgesamt haben wir also die folgende Kette von Inklusionen

$$R^{n+1} \subseteq R \circ T \subseteq T \circ T \subseteq T.$$

Damit folgt $R^{n+1} \subseteq T$ und der Induktions-Beweis ist abgeschlossen.

Wir zeigen nun, dass $R^+ \subseteq T$ ist. Sei $\langle x, y \rangle \in R^+$. Nach Definition von R^+ muss es dann eine positive natürliche Zahl n geben, so dass $\langle x, y \rangle \in R^n$ ist. Wegen $R^n \subseteq T$ folgt daraus aber $\langle x, y \rangle \in T$ und damit ist auch der zweite Teil des Beweises abgeschlossen. \square

Beispiel: Es sei *Mensch* die Menge alle Menschen, die jemals gelebt haben. Wir definieren die Relation *Eltern* auf *M* indem wir setzen

$$\text{Eltern} := \{\langle x, y \rangle \in \text{Mensch}^2 \mid x \text{ ist Vater von } y \text{ oder } x \text{ ist Mutter von } y\}$$

Dann besteht der transitive Abschluß der Relation *Eltern* aus allen Paaren $\langle x, y \rangle$, für die x ein Vorfahre von y ist:

$$\text{Eltern}^+ = \{\langle x, y \rangle \in \text{Mensch}^2 \mid x \text{ ist Vorfahre von } y\}.$$

Beispiel: Es sei *F* die Menge aller Flughäfen. Wir definieren auf der Menge *F* eine Relation *D* durch

$$D := \{\langle x, y \rangle \in F \times F \mid \text{Es gibt einen Direktflug von } x \text{ nach } y\}.$$

D bezeichnet also die direkten Verbindungen. Die Relation D^2 ist dann definiert als

$$D^2 = \{\langle x, z \rangle \in F \times F \mid \exists z \in F : \langle x, y \rangle \in D \wedge \langle y, z \rangle \in D\}.$$

Das sind aber gerade die Paare $\langle x, z \rangle$, für die es einen Luftweg von x nach z gibt, der genau einen Zwischenstop enthält. Entsprechend enthält D^3 die Paare $\langle x, z \rangle$, für die man mit zwei Zwischenstops von x nach z kommt und allgemein enthält D^k die Paare $\langle x, z \rangle$, für die man mit $k - 1$ Zwischenstops von dem Flughafen x zu dem

Flughafen z kommt. Der transitive Abschluß von D enthält dann alle Paare $\langle x, y \rangle$, für die es überhaupt eine Möglichkeit gibt, auf dem Luftweg von x nach y zu kommen.

Aufgabe 2: Auf der Menge \mathbb{N} der natürlichen Zahlen wird die Relation R wie folgt definiert:

$$R = \{\langle k, k+1 \rangle \mid k \in \mathbb{N}\}.$$

Berechnen Sie die folgenden Relationen:

1. R^2 ,
2. R^3 ,
3. R^n für beliebige $n \in \mathbb{N}$ mit $n \geq 1$,
4. R^+ .

Definition 10 Eine Relation $R \subseteq M \times M$ ist eine Äquivalenz-Relation auf M genau dann, wenn folgende Bedingungen erfüllt sind:

1. R ist reflexiv auf M ,
2. R ist symmetrisch und
3. R ist transitiv.

Der Begriff der Äquivalenz-Relationen verallgemeinert den Begriff der Gleichheit, denn ein triviales Beispiel für eine Äquivalenz-Relation auf M ist die Relation id_M . Als nicht-triviales Beispiel betrachten wir die Menge \mathbb{Z} der ganzen Zahlen zusammen mit der Relation \approx_n , die wir für natürliche Zahlen $n \neq 0$ wie folgt definieren:

$$\approx_n := \{\langle x, y \rangle \in \mathbb{Z}^2 \mid \exists k \in \mathbb{Z} : k \cdot n = x - y\}$$

Für zwei Zahlen $x, y \in \mathbb{Z}$ gilt also $x \approx_n y$ genau dann, wenn x und y beim Teilen durch n den gleichen Rest ergeben. Wir zeigen, dass die Relation \approx_n für $n \neq 0$ eine Äquivalenz-Relation auf \mathbb{Z} definiert.

1. Um zu zeigen, dass \approx_n reflexiv ist müssen wir nachweisen, dass für alle $x \in \mathbb{Z}$ gilt $\langle x, x \rangle \in \approx_n$. Nach Definition von \approx_n ist dies äquivalent zu

$$\langle x, x \rangle \in \{\langle x, y \rangle \in \mathbb{Z}^2 \mid \exists k \in \mathbb{Z} : k \cdot n = x - y\}.$$

Dies ist offenbar gleichbedeutend mit

$$\exists k \in \mathbb{Z} : k \cdot n = x - x.$$

Offenbar erfüllt $k = 0$ diese Gleichung, denn es gilt:

$$0 \cdot n = 0 = x - x.$$

Damit ist die Reflexivität nachgewiesen.

2. Um die Symmetrie von \approx_n nachzuweisen nehmen wir an, dass $\langle x, y \rangle \in \approx_n$ ist. Dann gibt es also ein $k \in \mathbb{Z}$, so dass

$$k \cdot n = x - y$$

gilt. Daraus folgt sofort

$$(-k) \cdot n = y - x.$$

Das zeigt aber, dass $\langle y, x \rangle \in \approx_n$ ist und damit ist die Symmetrie nachgewiesen.

3. Zum Nachweis der Transitivität von \approx nehmen wir an, dass sowohl $\langle x, y \rangle \in \approx_n$ als auch $\langle y, z \rangle \in \approx_n$ gelten. Dann gibt es also $k_1, k_2 \in \mathbb{Z}$ so dass

$$k_1 \cdot n = x - y \quad \text{und} \quad k_2 \cdot n = y - z$$

gelten. Addieren wir diese beiden Gleichungen, so sehen wir

$$(k_1 + k_2) \cdot n = x - z.$$

Mit $k_3 := k_1 + k_2$ gilt also $k_3 \cdot n = x - z$ und damit haben wir $\langle x, z \rangle \in \approx_n$ nachgewiesen und die Transitivität von \approx_n gezeigt. \square

Satz 11 *Es seien M und N Mengen und*

$$f : M \rightarrow N$$

sei eine Funktion. Definieren wir die Relation $R \subseteq M \times M$ als

$$R := \{ \langle x, y \rangle \in M \times M \mid f(x) = f(y) \}$$

so ist R eine Äquivalenz-Relation.

Beweis: Es weisen Reflexivität, Symmetrie und Transitivität von R nach:

1. R ist reflexiv, denn es gilt

$$\forall x \in M : f(x) = f(x).$$

Daraus folgt sofort

$$\forall x \in M : \langle x, x \rangle \in R.$$

2. Um die Symmetrie von R nachzuweisen, müssen wir

$$\forall x, y \in M : (\langle x, y \rangle \in R \rightarrow \langle y, x \rangle \in R)$$

zeigen. Sei also $\langle x, y \rangle \in R$. Dann gilt nach Definition von R

$$f(x) = f(y).$$

Daraus folgt sofort

$$f(y) = f(x)$$

und nach Definition von R ist das äquivalent zu

$$\langle y, x \rangle \in R.$$

3. Um die Transitivität von R nachzuweisen, müssen wir

$$\forall x, y, z \in M : (\langle x, y \rangle \in R \wedge \langle y, z \rangle \in R \rightarrow \langle x, z \rangle \in R)$$

zeigen. Gelte also

$$\langle x, y \rangle \in R \wedge \langle y, z \rangle \in R.$$

Nach Definition von R heißt das

$$f(x) = f(y) \wedge f(y) = f(z).$$

Daraus folgt sofort

$$f(x) = f(z).$$

Nach Definition der Relation R haben wir also

$$\langle x, z \rangle \in R.$$

\square

Beispiel: Es sei M die Menge aller Menschen und S sei die Menge aller Staaten. Nehmen wir zur Vereinfachung an, dass jeder Mensch genau eine Staatsbürgerschaft hat, so können wir eine Funktion

$$sb : M \rightarrow S$$

definieren, die jedem Menschen x seine Staatsbürgerschaft $sb(x)$ zuordnet. Bei der durch diese Funktion definierten Äquivalenz-Relation sind dann alle die Menschen äquivalent, welche die selbe Staatsbürgerschaft haben.

Definition 12 (Äquivalenz-Klasse) Ist R eine Äquivalenz-Relation auf M so definieren wir für alle $x \in M$ die Menge $[x]_R$ durch

$$[x]_R := \{y \in M \mid x R y\}. \quad (\text{Wir schreiben hier } xRy \text{ als Abkürzung für } \langle x, y \rangle \in R.)$$

Die Menge $[x]_R$ bezeichnen wir als die von x erzeugte Äquivalenz-Klasse.

Satz 13 Ist $R \subseteq M \times M$ eine Äquivalenz-Relation, so gilt:

1. $\forall x \in M: x \in [x]_R$
2. $\forall x, y \in M: (x R y \rightarrow [x]_R = [y]_R)$
3. $\forall x, y \in M: (\neg x R y \rightarrow [x]_R \cap [y]_R = \emptyset)$

Bemerkung: Da für $x, y \in M$ entweder $x R y$ oder $\neg(x R y)$ gilt, zeigen die letzten beiden Eigenschaften, dass zwei Äquivalenz-Klassen entweder gleich, oder aber disjunkt sind:

$$\forall x, y \in M: ([x]_R = [y]_R \vee [x]_R \cap [y]_R = \emptyset).$$

Beweis: Wir weisen nun die Gültigkeit der oben behaupteten Eigenschaften nach.

1. Wir haben $x \in [x]_R$ genau dann, wenn $x \in \{y \in M \mid x R y\}$ gilt und letzteres ist äquivalent zu $x R x$. Dies folgt unmittelbar aus der Reflexivität der Äquivalenz-Relation.
2. Sei nun $x R y$. Um $[x]_R = [y]_R$ nachzuweisen zeigen wir $[x]_R \subseteq [y]_R$ und $[y]_R \subseteq [x]_R$.

Sei also $u \in [x]_R$. Dann gilt $x R u$. Aus der Voraussetzung $x R y$ folgt wegen der Symmetrie der Relation R , dass auch $y R x$ gilt. Aus $y R x$ und $x R u$ folgt wegen der Transitivität der Relation R , dass $y R u$ gilt. Nach der Definition der Menge $[y]_R$ folgt damit $u \in [y]_R$. Damit ist $[x]_R \subseteq [y]_R$ nachgewiesen.

Um $[y]_R \subseteq [x]_R$ zu zeigen nehmen wir an, $u \in [y]_R$. Dann gilt $y R u$. Aus der Voraussetzung $x R y$ und $y R u$ folgt wegen der Transitivität der Relation R sofort $x R u$. Dann gilt aber $u \in [x]_R$ und damit ist auch die Inklusion $[y]_R \subseteq [x]_R$ nachgewiesen.

3. Sei nun $\neg(x R y)$ vorausgesetzt. Um nachzuweisen, dass $[x]_R \cap [y]_R = \emptyset$ ist nehmen wir an, dass es ein $z \in [x]_R \cap [y]_R$ gibt. Aus dieser Annahme werden wir einen Widerspruch zu der Voraussetzung $\neg(x R y)$ herleiten. Sei also $z \in [x]_R$ und $z \in [y]_R$. Nach Definition der Äquivalenz-Klassen $[x]_R$ und $[y]_R$ gilt dann

$$x R z \quad \text{und} \quad y R z.$$

Aufgrund der Symmetrie von R können wir $y R z$ umdrehen und haben dann

$$x R z \quad \text{und} \quad z R y.$$

Aus der Transitivität der Äquivalenz-Relation R folgt jetzt $x R y$. Dies steht aber im Widerspruch zu der Voraussetzung $\neg(x R y)$. Damit ist die Annahme, dass es ein $z \in [x]_R \cap [y]_R$ gibt, widerlegt. Folglich ist die Menge $[x]_R \cap [y]_R$ leer. \square

Beispiel: Setzen wir das letzte Beispiel fort, in dem alle die Menschen äquivalent waren, die die selbe Staatsbürgerschaft haben, so finden wir, dass die Äquivalenz-Klassen, die von dieser Äquivalenz-Relation erzeugt werden, gerade aus den Menschen besteht, die die selbe Staatsbürgerschaft besitzen.

Definition 14 (Partition) Ist $\mathcal{P} \subseteq 2^M$ eine Menge von Teilmengen von M , so sagen wir, dass \mathcal{P} eine Partition von M ist, falls folgende Eigenschaften gelten:

1. Vollständigkeits-Eigenschaft

$$\forall x \in M: \exists K \in \mathcal{P}: x \in K,$$

jedes Element aus M findet sich in einer Menge aus \mathcal{P} wieder.

2. Separations-Eigenschaft

$$\forall K, L \in \mathcal{P} : K \cap L = \emptyset \vee K = L,$$

zwei Mengen aus \mathcal{P} sind entweder disjunkt oder identisch.

Gelegentlich wird eine Partition einer Menge M auch als Zerlegung von M bezeichnet.

Bemerkung: Der letzte Satz hat gezeigt, dass für jede Äquivalenz-Relation R auf einer Menge M die Menge der Äquivalenz-Klassen

$$\{[x]_R \mid x \in M\}$$

eine Partition der Menge M darstellt. Der nächste Satz zeigt, dass auch die Umkehrung gilt, denn aus jeder Partition einer Menge läßt sich eine Äquivalenz-Relation erzeugen.

Satz 15 Es sei M eine Menge und $\mathcal{P} \subseteq 2^M$ eine Partition von M . Definieren wir die Relation R durch

$$R := \{\langle x, y \rangle \in M \times M \mid \exists K \in \mathcal{P} : (x \in K \wedge y \in K)\},$$

so ist R eine Äquivalenz-Relation auf M .

Beweis: Wir haben zu zeigen dass die Relation R reflexiv, symmetrisch und transitiv ist.

1. Reflexivität: Zu zeigen ist

$$\forall x \in M : x R x.$$

Das ist nach Definition der Relation R äquivalent zu der Formel

$$\forall x \in M : \exists K \in \mathcal{P} : (x \in K \wedge x \in K)$$

Das können wir sofort zu der Formel

$$\forall x \in M : \exists K \in \mathcal{P} : x \in K$$

vereinfachen. Diese Formel ist nichts anderes als die Vollständigkeit der Partition \mathcal{P} .

2. Symmetrie: Zu zeigen ist

$$\forall x, y \in M : (x R y \rightarrow y R x).$$

Wir nehmen also an, dass

$$x R y$$

gilt. Nach Definition der Relation R ist das äquivalent zu

$$\exists K \in \mathcal{P} : (x \in K \wedge y \in K).$$

Diese Formel ist offenbar äquivalent zu

$$\exists K \in \mathcal{P} : (y \in K \wedge x \in K)$$

und nach Definition der Relation R folgt nun

$$y R x.$$

3. Transitivität: Zu zeigen ist

$$\forall x, y, z \in M : (x R y \wedge y R z \rightarrow x R z).$$

Wir nehmen also an, dass

$$x R y \wedge y R z$$

gilt. Das ist nach Definition der Relation R äquivalent zu

$$\exists K \in \mathcal{P} : (x \in K \wedge y \in K) \wedge \exists L \in \mathcal{P} : (y \in L \wedge z \in L).$$

Dann gibt es aber offenbar zwei Mengen $K, L \in \mathcal{P}$, so dass

$$x \in K \wedge y \in K \cap L \wedge z \in L$$

gilt. Damit ist $K \cap L \neq \emptyset$ und aus der Separations-Eigenschaft der Partition \mathcal{P} folgt

$$K = L.$$

Damit haben wir

$$\exists K \in \mathcal{P} : (x \in K \wedge z \in K)$$

gezeigt und nach Definition der Relation R heißt das

$$x R z.$$

□

Partielle Ordnung, Totale Ordnung

Eine Relation $R \subseteq M \times M$ ist eine *partielle Ordnung* (im Sinne von \leq) auf M genau dann, wenn R

1. reflexiv,
2. anti-symmetrisch und
3. transitiv ist.

Die Relation ist darüber hinaus eine *totale Ordnung* auf M , wenn gilt:

$$\forall x \in M : \forall y \in M : (x R y \vee y R x).$$

Beispiel: Die Teilbarkeitsrelation div kann auf den natürlichen Zahlen wie folgt definiert werden

$$\text{div} := \{ \langle x, y \rangle \in \mathbb{N} \times \mathbb{N} \mid \exists k \in \mathbb{N} : k \cdot x = y \}.$$

Wir zeigen dass diese Relation eine partielle Ordnung auf \mathbb{N} ist und weisen dazu Reflexivität, Anti-Symmetrie und Transitivität nach.

1. Reflexivität: Zu zeigen ist

$$\forall x \in \mathbb{N} : x \text{ div } x.$$

Nach Definition der Relation div ist das äquivalent zu

$$\exists k \in \mathbb{N} : k \cdot x = x$$

Setzen wir $k = 1$, so gilt sicher $k \cdot x = x$ und damit ist die Reflexivität gezeigt.

2. Anti-Symmetrie: Zu zeigen ist

$$\forall x, y \in \mathbb{N} : (x \text{ div } y \wedge y \text{ div } x \rightarrow x = y)$$

Wir nehmen also an, dass

$$x \text{ div } y \wedge y \text{ div } x$$

gilt (und werden $x = y$ zeigen). Nach Definition der Relation div ist die Annahme äquivalent zu

$$(\exists k_1 \in \mathbb{N} : k_1 \cdot x = y) \wedge (\exists k_2 \in \mathbb{N} : k_2 \cdot y = x)$$

Also gibt es natürliche Zahlen k_1 und k_2 , so dass

$$k_1 \cdot x = y \wedge k_2 \cdot y = x$$

gilt. Setzen wir diese Gleichungen ineinander ein, so erhalten wir

$$k_1 \cdot k_2 \cdot y = y \quad \text{und} \quad k_2 \cdot k_1 \cdot x = x.$$

Dann muss aber

$$k_1 \cdot k_2 = 1 \vee (x = 0 \wedge y = 0)$$

gelten. Da aus $k_1 \cdot k_2 = 1$ sofort $k_1 = 1$ und $k_2 = 1$ folgt haben wir wegen der ursprünglichen Gleichungen $k_1 \cdot x = y$ und $k_2 \cdot y = x$ in jedem Fall $x = y$.

3. Transitivität: Zu zeigen ist

$$\forall x, y, z \in \mathbb{N} : (x \text{ div } y \wedge y \text{ div } z \rightarrow x \text{ div } z)$$

Wir nehmen also an, dass

$$x \text{ div } y \wedge y \text{ div } z$$

gilt (und werden $x \text{ div } z$ zeigen). Nach Definition der Relation div ist die Annahme äquivalent zu

$$(\exists k_1 \in \mathbb{N} : k_1 \cdot x = y) \wedge (\exists k_2 \in \mathbb{N} : k_2 \cdot y = z)$$

Also gibt es natürliche Zahlen k_1 und k_2 , so dass

$$k_1 \cdot x = y \wedge k_2 \cdot y = z$$

gilt. Setzen wir die erste Gleichung in die zweite ein, so erhalten wir

$$k_2 \cdot k_1 \cdot x = z.$$

Setzen wir $k_3 := k_2 \cdot k_1$, so haben wir also $k_3 \cdot x = z$ und das zeigt

$$x \text{ div } z.$$

Die Relation div ist keine totale Ordnung, denn beispielsweise gilt weder $2 \text{ div } 3$ noch $3 \text{ div } 2$. ◇

Aufgabe 3: Auf der Menge der ganzen Zahlen \mathbb{N} definieren wir die Relation \leq wie folgt:

$$\leq := \{ \langle x, y \rangle \in \mathbb{N} \times \mathbb{N} \mid \exists k \in \mathbb{N} : x + k = y \}.$$

Zeigen Sie, dass die Relation \leq eine totale Ordnung auf \mathbb{N} ist. ◇

Aufgabe 4: Auf der Potenz-Menge der natürlichen Zahlen definieren wir die Relation \subseteq als

$$\subseteq := \{ \langle A, B \rangle \in 2^{\mathbb{N}} \times 2^{\mathbb{N}} \mid \exists C \in 2^{\mathbb{N}} : A \cup C = B \}$$

Zeigen Sie, dass die Relation \subseteq auf $2^{\mathbb{N}}$ zwar eine partielle, aber keine totale Ordnung ist. ◇

Wir schließen damit den theoretischen Teil unseres Ausflugs in die Mengenlehre und verweisen für weitere Details auf die Literatur [Lip98].

Kapitel 3

Die Programmier-Sprache SetlX

Die im letzten Kapitel vorgestellten Begriffs-Bildungen aus der Mengenlehre bereiten erfahrungsgemäß dem Anfänger aufgrund ihrer Abstraktheit gewisse Schwierigkeiten. Um diese Begriffe vertrauter werden zu lassen, stelle ich daher nun eine Programmier-Sprache vor, die mit diesen Begriffen arbeitet. Dies ist die Sprache SETLX. Diese Sprache basiert auf der Ende der siebziger Jahre von Jacob T. Schwartz eingeführten Sprache SETL [SDSD86]. Die Sprache SETLX lehnt sich in ihrer Syntax stark an die Programmiersprache C an, ist vom Konzept her aber als Derivat von SETL zu sehen. Sie finden auf der Webseite

<http://wwwlehre.dhbw-stuttgart.de/stroetmann/SetlX/setlX.php>

eine Anleitung zur Installation von SETLX.

3.1 Einführende Beispiele

Wir wollen in diesem Abschnitt die Sprache SETLX an Hand einiger einfacher Beispiele vorstellen, bevor wir dann in den folgenden Abschnitten auf die Details eingehen.

```
1  -----setlX-----v2.2.0--
2
3  Welcome to the setlX interpreter!
4
5  Open Source Software from http://setlX.randoom.org/
6  (c) 2011-2013 by Herrmann, Tom
7
8  You can display some helpful information by using '--help' as parameter when
9  launching this program.
10
11 Interactive-Mode:
12   The 'exit;' statement terminates the interpreter.
13
14 -----Interactive=Mode-----
15
16 =>
```

Abbildung 3.1: SETLX-Interpreter nach dem Start.

Die Sprache SETLX ist eine interaktive Sprache, Sie können diese Sprache also unmittelbar über einen Interpreter aufrufen. Falls Sie SETLX auf Ihrem Rechner installiert haben, können Sie den Befehl

```
setlx
```

in einer Kommando-Zeile eingeben. Anschließend meldet sich der Interpreter dann wie in Abbildung 3.1 auf Seite 37 gezeigt. Die Zeichenfolge “=>” ist der Prompt, der Ihnen signalisiert, dass der Interpreter auf eine Eingabe wartet. Geben Sie dort den Text

```
1 + 2;
```

ein und drücken anschließend auf die Eingabe-Taste, so erhalten Sie die folgende Ausgabe:

```
1  Hallo
2  ~< Result: 3 >~
3
4  =>
```

Hier hat der Interpreter die Summe $1 + 2$ berechnet und das Ergebnis ausgegeben. Es gibt auch eine Funktion, mit deren Hilfe Sie Werte ausdrucken können. Diese Funktion heißt `print`. Wenn Sie im Interpreter

```
print("Hallo");
```

eingeben und anschließend auf die Eingabe-Taste drücken, so erhalten Sie die folgende Ausgabe:

```
1  Hallo
2  ~< Result: om >~
3
4  =>
```

Hier hat der Interpreter zunächst den Befehl `print("Hallo")` ausgeführt und dabei den Text “Hallo” ausgegeben. In der Zeile darunter wird der Wert des zuletzt ausgegebenen Ausdrucks angezeigt. Da die Funktion `print()` kein Ergebnis berechnet, ist der Rückgabe-Wert undefiniert. Ein undefinierter Wert wird in SETLX mit Ω bezeichnet, was in der Ausgabe durch den String “om” dargestellt wird.

Die Funktion `print()` akzeptiert beliebig viele Argumente. Wenn Sie beispielsweise das Ergebnis der Rechnung $36 \cdot 37 / 2$ ausgeben wollen, so können Sie dies über den Befehl

```
print("36 * 37 / 2 = ", 36 * 37 / 2);
```

erreichen. Wenn Sie nur an der Auswertung dieses Ausdrucks interessiert sind, so können Sie diesen Ausdruck auch unmittelbar hinter dem Prompt eingeben und mit einem Semikolon “;” abschließen. Wenn Sie nun die Eingabe-Taste betätigen, wird der Ausdruck ausgewertet und das Ergebnis angezeigt.

Der SETLX-Interpreter läßt sich nicht nur interaktiv betreiben, sondern er kann auch vollständige Programme ausführen. Speichern wir das in Abbildung 3.2 gezeigte Programm in einer Datei mit dem Namen “sum.stlx” ab, so können wir in der Kommando-Zeile den Befehl

```
setlx sum.stlx
```

eingeben. Dann wird zunächst der Text “Type a natural number ...” gefolgt von einem Doppelpunkt als Prompt ausgegeben. Geben wir nun eine Zahl n ein und betätigen die Eingabe-Taste, so wird als Ergebnis die Summe $\sum_{i=1}^n i$ ausgegeben.

Wir diskutieren nun das in Abbildung 3.2 auf Seite 39 gezeigte Programm Zeile für Zeile. Die Zeilen-Nummern in dieser und den folgenden Abbildungen von SETLX-Programmen sind nicht Bestandteil der Programme sondern wurden hinzugefügt um in den Diskussionen dieser Programme besser auf einzelne Zeilen Bezug nehmen zu können.

```

1  // This program reads a number n and computes the sum 1 + 2 + ... + n.
2  n := read("Type a natural number and press return: ");
3  s := +/ { 1 .. n };
4  print("The sum 1 + 2 + ... + ", n, " is equal to ", s, ".");

```

Abbildung 3.2: Ein einfaches Programm zur Berechnung der Summe $\sum_{i=1}^n i$.

1. Die erste Zeile enthält einen Kommentar. In SETLX werden Kommentare durch den String “//” eingeleitet. Aller Text zwischen diesem String und dem Ende der Zeile wird von dem SETLX-Compiler ignoriert. Neben einzeiligen Kommentaren unterstützt SETLX auch mehrzeilige Kommentare, die wie in der Sprache C durch die Strings “/*” und “*/” begrenzt werden.
2. Die zweite Zeile enthält eine Zuweisung. Die Funktion `read()` gibt zunächst den Text aus, der den Benutzer zur Eingabe einer Zahl auffordert und liest dann den vom Benutzer eingegebenen String. Falls der Benutzer eine Zahl eingibt, wird dies erkannt und die eingegebene Zahl wird Hilfe des Zuweisungs-Operators “:=” der Variablen `n` zugewiesen. An dieser Stelle gibt es zwei wichtige Unterschiede zur Syntax der Sprache C:
 - (a) **SetlX** verwendet den Zuweisungs-Operator “:=”, während die Programmiersprache C den String “=” als Zuweisungs-Operator benutzt.
 - (b) Alle Variablen müssen in der Sprache SETLX mit einem kleinen Buchstaben beginnen. In C können Variablen auch mit einem großen Buchstaben beginnen. Das selbe gilt für die Namen von Funktionen.

Im Gegensatz zu der Sprache C ist die Sprache SETLX *ungetypt*. Daher ist es weder notwendig noch möglich, die Variable `n` zu deklarieren. Würde der Benutzer an Stelle einer Zahl einen String eingeben, so würde das Programm später mit einer Fehlermeldung abbrechen.

3. Die dritte Zeile zeigt zunächst, wie sich Mengen als Aufzählungen definieren lassen. Sind a und b ganze Zahlen mit $a < b$, so berechnet der Ausdruck

$$\{ a \dots b \}$$

die Menge

$$\{x \in \mathbb{Z} \mid a \leq x \wedge x \leq b\}.$$

Der Operator “+/” berechnet dann die Summe aller Elemente der Menge

$$\{i \in \mathbb{N} \mid 1 \leq i \wedge i \leq n\}$$

Das ist natürlich genau die Summe

$$1 + 2 + \dots + n = \sum_{i=1}^n i.$$

Diese Summe wird der Variablen `s` zugewiesen.

4. In der letzten Zeile wird diese Summe ausgegeben.

Als nächstes betrachten wir das in Abbildung 3.3 auf Seite 40 gezeigte Programm `sum-recursive.stlx`, das die Summe $\sum_{i=0}^n i$ mit Hilfe einer Rekursion berechnet.

1. Zeile 1 bis Zeile 7 enthalten die Definition der Prozedur `sum`. Die Definition einer Prozedur wird in SETLX durch das Schlüsselwort “**procedure**” eingeleitet. Hinter diesem Schlüsselwort folgt zunächst eine öffnende Klammer “(”, dann eine Liste von Argumenten, welche durch “,” voneinander getrennt sind, und danach eine schließenden Klammer “)”. Darauf folgt der Rumpf der Prozedur, der, genau wie in der Sprache C, durch die Klammern “{” und “}” begrenzt wird. Im Allgemeinen besteht der Rumpf aus einer Liste

```

1  sum := procedure(n) {
2      if (n == 0) {
3          return 0;
4      } else {
5          return sum(n-1) + n;
6      }
7  };
8
9  n      := read(("Zahl eingeben: "));
10 total := sum(n);
11 print("Sum 0 + 1 + 2 + ... + ", n, " = ", total);

```

Abbildung 3.3: Ein rekursives Programm zur Berechnung der Summe $\sum_{i=0}^n i$.

von Kommandos. In unserem Fall haben wir hier nur ein einziges Kommando. Dieses Kommando ist allerdings ein zusammengesetztes Kommando und zwar eine Fallunterscheidung. Die allgemeine Form einer Fallunterscheidung ist wie folgt:

```

1      if ( test ) {
2          body1
3      } else {
4          body2
5      }

```

Die Semantik der Fallunterscheidung ist wie folgt:

- Zunächst wird der Ausdruck *test* ausgewertet. Bei der Auswertung muß sich entweder der Wert “true” oder “false” ergeben.
- Falls sich “true” ergibt, werden anschließend die Kommandos in *body₁* ausgeführt. Dabei ist *body₁* eine Liste von Kommandos.
- Andernfalls werden die Kommandos in der Liste *body₂* ausgeführt.

Beachten Sie hier die folgenden beiden Unterschiede zur Sprache C:

- Bei dem `if`-Befehl müssen Sie auch dann geschweifte Klammern verwenden, wenn *body₁* und *body₂* nur aus einem einzigen Befehl bestehen.
 - Die Definition der Prozedur muss durch ein Semikolon abgeschlossen werden.
- Nach der Definition der Prozedur `sum` wird in Zeile 9 ein Wert in die Variable `n` eingelesen.
 - Dann wird in Zeile 10 für den eben eingelesenen Wert von `n` die oben definierte Prozedur `sum` aufgerufen. Zusätzlich enthält Zeile 10 eine Zuweisung: Der Wert, den der Prozedur-Aufruf `sum(n)` zurück liefert, wird in die Variable `total`, die auf der linken Seite des *Zuweisungs-Operators* “:=” steht, geschrieben.
 - Anschließend wird das berechnete Ergebnis durch einen `print`-Befehl ausgegeben.

Die Prozedur `sum` in dem obigen Beispiel ist *rekursiv*, d.h. sie ruft sich selber auf. Die Logik, die hinter der Implementierung steht, läßt sich am einfachsten durch die beiden folgenden bedingten Gleichungen erfassen:

- $sum(0) = 0$,

$$2. \ n > 0 \rightarrow \text{sum}(n) = \text{sum}(n - 1) + n.$$

Die Korrektheit dieser Gleichungen wird unmittelbar klar, wenn wir für $\text{sum}(n)$ die Definition

$$\text{sum}(n) = \sum_{i=0}^n i$$

einsetzen, denn offenbar gilt:

1. $\text{sum}(0) = \sum_{i=0}^0 i = 0,$
2. $\text{sum}(n) = \sum_{i=0}^n i = \left(\sum_{i=0}^{n-1} i \right) + n = \text{sum}(n - 1) + n.$

Die erste Gleichung behandelt den Fall, dass die Prozedur sich nicht selbst aufruft. Einen solchen Fall muss es in jeder rekursiv definierten Prozedur geben, denn sonst würde die Prozedur in einer Endlos-Schleife stecken bleiben.

3.2 Darstellung von Mengen

Der wichtigste Unterschied zwischen der Sprache SETLX und der Sprache C besteht darin, dass SETLX die Verwendung von Mengen und Listen unmittelbar unterstützt. Um zu zeigen, wie wir in SETLX mit Mengen umgehen können, zeigen wir ein einfaches Programm, das Vereinigung, Schnitt und Differenz zweier Mengen berechnet. Abbildung 3.4 zeigt die Datei `simple.stlx`. Das in dieser Abbildung gezeigte Programm zeigt die Verwendung der elementaren Mengen-Operatoren in SETLX.

```

1  a := { 1, 2, 3 };
2  b := { 2, 3, 4 };
3  // Berechnung der Vereinigungs-Menge a ∪ b
4  c := a + b;
5  print(a, " + ", b, " = ", c);
6  // Berechnung der Schnitt-Menge      a ∩ b
7  c := a * b;
8  print(a, " * ", b, " = ", c);
9  // Berechnung der Mengen-Differenz   a \ b
10 c := a - b;
11 print(a, " - ", b, " = ", c);
12 // Berechnung der Potenz-Menge      2a
13 c := pow(a);
14 print("pow(", a, ") = ", c);
15 // Überprüfung einer Teilmengen-Beziehung   a ⊆ b
16 print("(", a, " <= ", b, ") = ", (a <= b));
```

Abbildung 3.4: Berechnung von \cup , \cap , \setminus und Potenz-Menge

In Zeile 1 und 2 sehen wir, dass wir Mengen ganz einfach durch explizite Aufzählung ihrer Argumente angeben können. In den Zeilen 4, 7 und 10 berechnen wir dann nacheinander Vereinigung, Schnitt und Differenz dieser Mengen. Hier ist zu beachten, dass dafür in SETLX die Operatoren “+”, “*”, “−” verwendet werden. In Zeile 13 berechnen wir die Potenz-Menge mit Hilfe der Funktion `pow()`. Schließlich überprüfen wir in Zeile 17 mit dem Operator “<=”, ob `a` eine Teilmenge von `b` ist. Führen wir dieses Programm aus, so erhalten wir die folgende Ausgabe:

```

{1, 2, 3} + {2, 3, 4} = {1, 2, 3, 4}
{1, 2, 3} * {2, 3, 4} = {2, 3}
{1, 2, 3} - {2, 3, 4} = {1}
pow({1, 2, 3}) = {}, {1}, {1, 2}, {1, 2, 3}, {1, 3}, {2}, {2, 3}, {3}
({1, 2, 3} <= {2, 3, 4}) = false

```

Um interessantere Programme zeigen zu können, stellen wir jetzt weitere Möglichkeiten vor, mit denen wir in SETLX Mengen definieren können.

Definition von Mengen durch arithmetische Aufzählung

In dem letzten Beispiel hatten wir Mengen durch explizite Aufzählung definiert. Das ist bei großen Mengen viel zu mühsam. Eine Alternative ist daher die Definition einer Menge durch eine *arithmetische Aufzählung*. Wir betrachten zunächst ein Beispiel:

```
a := { 1 .. 100 };
```

Die Menge, die hier der Variablen `a` zugewiesen wird, ist die Menge aller natürlichen Zahlen von 1 bis 100. Die allgemeine Form einer solchen Definition ist

```
a := { start .. stop };
```

Mit dieser Definition würde `a` die Menge aller ganzen Zahlen von `start` bis `stop` zugewiesen, formal gilt

$$a = \{n \in \mathbb{Z} \mid start \leq n \wedge n \leq stop\}.$$

Es gibt noch eine Variante der arithmetischen Aufzählung, die wir ebenfalls durch ein Beispiel einführen.

```
a := { 1, 3 .. 100 };
```

Die Menge, die hier der Variablen `a` zugewiesen wird, ist die Menge aller ungeraden natürlichen Zahlen von 1 bis 100. Die Zahl 100 liegt natürlich nicht in dieser Menge, denn sie ist ja gerade. Die allgemeine Form einer solchen Definition ist

```
a := { start, second .. stop }
```

Definieren wir `step = second - start` und ist `step` positiv, so läßt sich diese Menge formal wie folgt definieren:

$$a = \{start + n * step \mid n \in \mathbb{N} \wedge start + n * step \leq stop\}.$$

Beachten Sie, dass `stop` nicht unbedingt ein Element der Menge

```
a := { start, second .. stop }
```

ist. Beispielsweise gilt

$$\{ 1, 3 .. 6 \} = \{ 1, 3, 5 \}.$$

Definition von Mengen durch Iteratoren

Eine weitere Möglichkeit, Mengen zu definieren, ist durch die Verwendung von *Iteratoren* gegeben. Wir geben zunächst ein einfaches Beispiel:

```
p := { n * m : n in {2..10}, m in {2..10} };
```

Nach dieser Zuweisung enthält `p` die Menge aller *nicht-trivialen* Produkte, deren Faktoren ≤ 10 sind. (Ein Produkt der Form $a \cdot b$ heißt dabei *trivial* genau dann, wenn einer der Faktoren a oder b den Wert 1 hat.) In der Schreibweise der Mathematik gilt für die oben definierte Menge `p`:

$$p = \{n \cdot m \mid n \in \mathbb{N} \wedge m \in \mathbb{N} \wedge 2 \leq n \wedge 2 \leq m \wedge n \leq 10 \wedge m \leq 10\}.$$

Wie ausdrucksstark Iteratoren sind, läßt sich an dem Programm `primes-difference.stlx` erkennen, das in Abbildung 3.5 auf Seite 43 gezeigt ist. Das Programm berechnet die Menge der Primzahlen bis zu einer vorgegebenen Größe n und ist so kurz wie eindrucksvoll. Die zugrunde liegende Idee ist, dass eine Zahl genau dann

eine Primzahl ist, wenn Sie von 1 verschieden ist und sich nicht als Produkt zweier von 1 verschiedener Zahlen schreiben läßt. Um also die Menge der Primzahlen kleiner gleich n zu berechnen, ziehen wir einfach von der Menge $\{2, \dots, n\}$ die Menge aller Produkte ab. Genau dies passiert in Zeile 2 des Programms.

```

1  n := 100;
2  primes := {2 .. n} - { p * q : p in {2..n}, q in {2..n} };
3  print(primes);

```

Abbildung 3.5: Programm zur Berechnung der Primzahlen bis n .

Die allgemeine Form der Definition eine Menge mit Iteratoren ist durch den Ausdruck

$$\{expr : x_1 \text{ in } S_1, \dots, x_n \text{ in } S_n\}$$

gegeben. Hierbei ist $expr$ ein Term, in dem die Variablen x_1 bis x_n auftreten. Weiterhin sind S_1 bis S_n Ausdrücke, die bei ihrer Auswertung Mengen ergeben. Die Ausdrücke $x_i \text{ in } S_i$ werden dabei als *Iteratoren* bezeichnet, weil die Variablen x_i über alle Werte der entsprechenden Menge S_i laufen (wir sagen auch: *iterieren*). Die mathematische Interpretation der obigen Menge ist dann durch

$$\{expr \mid x_1 \in S_1 \wedge \dots \wedge x_n \in S_n\}$$

gegeben. Die Definition einer Menge über Iteratoren entspricht der Definition einer Menge als Bild-Menge.

Es ist in SETLX auch möglich, das *Auswahl-Prinzip* zu verwenden. Dazu können wir Iteratoren mit einer Bedingung verknüpfen. Die allgemeine Syntax dafür ist:

$$\{expr : x_1 \text{ in } S_1, \dots, x_n \text{ in } S_n \mid cond\}.$$

Hierbei haben die Ausdrücke $expr$ und S_i die selbe Bedeutung wie oben und $cond$ ist ein Ausdruck, der von den Variablen x_1, \dots, x_n abhängt und dessen Auswertung entweder **true** oder **false** ergibt. Die mathematische Interpretation der obigen Menge ist dann

$$\{expr \mid x_1 \in S_1 \wedge \dots \wedge x_n \in S_n \wedge cond\}.$$

Wir geben ein konkretes Beispiel: Nach der Zuweisung

```
primes := { p : p in {2..100} | { x : x in {1..p} | p % x == 0 } == {1, p} };
```

enthält **primes** die Menge aller Primzahlen, die kleiner oder gleich 100 sind. Die der obigen Berechnung zugrunde liegende Idee besteht darin, dass eine Zahl genau dann eine Primzahl ist, wenn Sie nur durch 1 und sich selbst teilbar ist. Um festzustellen, ob eine Zahl p durch eine andere Zahl x teilbar ist, können wir in SETLX den Operator **%** verwenden: Der Ausdruck $p \% x$ berechnet den Rest, der übrig bleibt, wenn Sie die Zahl p durch x teilen. Eine Zahl p ist also genau dann durch eine andere Zahl x teilbar, wenn der Rest 0 ist, wenn also $p \% x = 0$ gilt. Damit liefert

$$\{ x : x \text{ in } \{1..p\} \mid p \% x == 0 \}$$

genau die Menge aller Teiler von p und p ist eine Primzahl, wenn diese Menge nur aus den beiden Zahlen 1 und p besteht. Das Programm aus der Datei **primes-slim.stlx**, das in Abbildung 3.6 auf Seite 44 gezeigt wird, benutzt diese Methode zur Berechnung der Primzahlen.

Zunächst haben wir in den Zeilen 1 bis 3 die Funktion **teiler(p)** definiert, die als Ergebnis die Menge der Teiler der Zahl p berechnet. Die Menge der Primzahlen ist dann die Menge der Zahlen p , die nur den Teiler 1 oder p haben.

3.3 Paare, Relationen und Funktionen

Das Paar $\langle x, y \rangle$ wird in SETLX in der Form $[x, y]$ dargestellt, die spitzen Klammern werden also durch eckige Klammern ersetzt. Wir hatten im letzten Kapitel gesehen, dass wir eine Menge von Paaren, die sowohl links-

```

1  teiler := procedure(p) {
2      return { t : t in {1..p} | p % t == 0 };
3  };
4  n      := 100;
5  primes := { p : p in {2..n} | teiler(p) == {1, p} };
6  print(primes);

```

Abbildung 3.6: Alternatives Programm zur Berechnung der Primzahlen.

total als auch rechts-eindeutig ist, auch als Funktion auffassen können. Ist R eine solche Menge und $x \in \text{dom}(R)$, so bezeichnet in SETLX der Ausdruck $R(x)$ das eindeutig bestimmte Element y , für das $\langle x, y \rangle \in R$ gilt. Das Programm `function.stlx` in Abbildung 3.7 auf Seite 44 zeigt dies konkret. Zusätzlich zeigt das Programm noch, dass in SETLX bei einer binären Relation $\text{dom}(R)$ als `domain(R)` und $\text{rng}(R)$ als `range(R)` geschrieben wird. Außerdem sehen wir in Zeile 2, dass wir den Wert einer funktionalen Relation durch eine Zuweisung ändern können.

```

1  q := { [n, n**2] : n in {1..10} };
2  q[5] := 7;
3  print( "q[3]   = ", q[3]       );
4  print( "q[5]   = ", q[5]       );
5  print( "dom(q) = ", domain(q) );
6  print( "rng(q) = ", range(q)  );
7  print( "q = ", q );

```

Abbildung 3.7: Rechnen mit binären Relationen.

Das Programm berechnet in Zeile 1 die binäre Relation q so, dass q die Funktion $x \mapsto x * x$ auf der Menge $\{n \in \mathbb{N} \mid 1 \leq n \wedge n \leq 10\}$ repräsentiert.

An dieser Stelle eine Warnung: In SETLX müssen alle Variablen mit einem kleinen Buchstaben beginnen! Normalerweise hätte ich die Relation q mit einem großen Q bezeichnen wollen, aber das geht nicht, denn alle Namen, die mit einem großen Buchstaben beginnen, sind zur Darstellung von *Termen* reserviert.

In Zeile 2 wird die Relation an der Stelle $x = 5$ so abgeändert, dass nun $q(5)$ den Wert 7 hat. Anschließend werden noch $\text{dom}(Q)$ und $\text{rng}(Q)$ berechnet. Das Programm liefert die folgende Ausgabe:

```

q[3]   = 9
q[5]   = 7
dom(q) = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
rng(q) = {1, 4, 7, 9, 16, 36, 49, 64, 81, 100}
q = {[1, 1], [2, 4], [3, 9], [4, 16], [5, 7], [6, 36], [7, 49], [8, 64],
     [9, 81], [10, 100]}

```

Es ist naheliegend zu fragen, was bei der Auswertung eines Ausdrucks der Form $R(x)$ passiert, wenn die Menge $\{y \mid \langle x, y \rangle \in R\}$ entweder leer ist oder aber aus mehr als einem Element besteht. Das Programm `buggy-function.stlx` in Abbildung 3.8 auf Seite 45 beantwortet diese Frage auf experimentellem Wege.

Falls die Menge $\{y \mid \langle x, y \rangle \in r\}$ entweder leer ist oder mehr als eine Element enthält, so ist der Ausdruck $r(x)$ undefiniert. Ein solcher Ausdruck wird in SETLX durch den String “om” dargestellt. Der Versuch, einen undefinierten Wert in eine Menge M einzufügen, ändert diese Menge nicht, es gibt aber auch keine Fehlermeldung. Deswegen wird in Zeile 4 für die Menge $\{r(1), r(2)\}$ einfach die leere Menge ausgegeben.

Will man das Auftreten von undefinierten Werten beim Auswerten einer binären Relation r vermeiden, so gibt es in SETLX die Möglichkeit, $r\{x\}$ statt $r(x)$ zu schreiben, die runden Klammern werden also durch geschweifte Klammern ersetzt. Der Ausdruck $r\{x\}$ ist für eine binäre Relation r wie folgt definiert:

```

1  r := { [1, 1], [1, 4], [3, 3] };
2  print( "r[1] = ", r[1] );
3  print( "r[2] = ", r[2] );
4  print( "{ r[1], r[2] } = ", { r[1], r[2] } );
5  print( "r{1} = ", r{1} );
6  print( "r{2} = ", r{2} );

```

Abbildung 3.8: Rechnen mit nicht-funktionalen binären Relationen.

$$r\{x\} := \{y \mid \langle x, y \rangle \in r\}$$

Daher liefert das Programm aus Abbildung 3.8 die folgende Ausgabe:

```

r[1] = om
r[2] = om
{ r[1], r[2] } = {}
r{1} = {1, 4}
r{2} = {}

```

3.4 Allgemeine Tupel

Auch beliebige n -Tupel lassen sich in SETLX darstellen. Diese können ganz analog zu Mengen definiert werden. Das geht denkbar einfach: Es müssen nur alle geschweiften Klammern der Form “{” und “}” durch die entsprechenden eckigen Klammern “[” und “]” ersetzt werden. Dann können Tupel durch explizite Aufzählung, arithmetische Aufzählung, Iteration und das Auswahlprinzip in der selben Weise wie Mengen gebildet werden. In SETLX werden Tupel in der Regel als *Listen* bezeichnet. Das Programm `primes-tuple.stlx` in Abbildung 3.9 zeigt, wie wir in SETLX mit Listen arbeiten können. Dieses Programm berechnet die Primzahlen nach dem selben Verfahren wie das Programm in Abbildung 3.6 auf Seite 44, benutzt aber Listen sowohl zur Darstellung der Menge der Primzahlen als auch zur Darstellung der Menge der Teiler.

```

1  teiler := procedure(p) {
2      return [ t : t in {1..p} | p % t == 0 ];
3  };
4
5  n := 100;
6  primes := [ p : p in [2 .. n] | teiler(p) == [1, p] ];
7  print(primes);

```

Abbildung 3.9: Berechnung der Primzahlen mit Tupeln.

3.5 Spezielle Funktionen und Operatoren auf Mengen

Das Programm `sort.stlx` in Abbildung 3.10 auf Seite 46 zeigt ein einfaches Verfahren um eine Liste von Zahlen zu sortieren. Der Ausdruck

```
max(s)
```

berechnet das größte Element der Liste s . Damit läuft die Variable n in dem Iterator

```
n in [1 .. max(s)]
```

dann von 1 bis zur größten in s auftretenden Zahl. Aufgrund der Bedingung “ n in s ” wird die Zahl n genau dann in die resultierende Liste eingefügt, wenn n ein Element der Liste s ist. Da der Iterator

```
“n in [1 .. max(s)]”
```

die Zahlen der Reihe nach aufzählt, ist das Ergebnis eine sortierte Liste, die genau die Zahlen enthält, die Elemente von s sind.

Offensichtlich ist der in der Prozedur `sort()` implementierte Algorithmus nicht sehr effizient. Wir werden später noch effizientere Algorithmen diskutieren.

```
1  sort := procedure(s) {
2      return [ n in [1 .. max(s)] | n in s ];
3  };
4  s := [ 13, 5, 7, 2, 4 ];
5  print( "sort( ", s, " ) = ", sort(s) );
```

Abbildung 3.10: Sortieren einer Menge.

Analog zu der Funktion `max()` gibt es noch die Funktion `min()` die das Minimum einer Menge oder Liste berechnet.

Weiterhin können die Operatoren “+” und “*” auf Mengen angewendet werden. Der Operator “+” berechnet die Summe aller Elemente einer Menge, während der Operator “*” das Produkt der Elemente berechnet. Ist die zu Grunde liegende Menge oder Liste leer, so geben diese Operatoren “+” als Ergebnis “om” zurück. Die Operatoren “+” und “*” können auch als binäre Operatoren verwendet werden. Ein Ausdruck der Form

$x +/ s$

gibt als Ergebnis den Wert x zurück, wenn die Menge s leer ist. Falls s nicht leer ist, liefert der obige Ausdruck die Summe der Elemente der Menge s . Ein Ausdruck der Form “ $x */ s$ ” funktioniert analog: Falls s leer ist, wird x zurück gegeben, sonst ist das Ergebnis das Produkt der Elemente aus s .

Als nächstes besprechen wir die Funktion “`from`”, mit dem wir ein (nicht näher spezifiziertes) Element aus einer Menge auswählen können. Die Syntax ist:

```
x := from(s);
```

Hierbei ist s eine Menge und x eine Variable. Wird diese Anweisung ausgeführt, so wird ein nicht näher spezifiziertes Element aus der Menge s entfernt. Dieses Element wird darüber hinaus der Variablen x zugewiesen. Falls s leer ist, so erhält x den undefinierten Wert “om” und s bleibt unverändert. Das Programm `from.stlx` in Abbildung 3.11 auf Seite 47 nutzt diese Anweisung um eine Menge elementweise auszugeben. Jedes Element wird dabei in einer eigenen Zeile ausgedruckt.

Neben der Funktion “`from`” gibt es noch die Funktion “`arb`”, die ein beliebiges Element aus einer Menge auswählt, die Menge selbst aber unverändert läßt. Nach den Befehlen

```
s := { 1, 2 };
x := arb(s);
print("x = ", x);
print("s = ", s);
```

erhalten wir die folgende Ausgabe:

```
x = 13
s = {2, 3, 5, 7, 13}
```

Weiterhin steht für Listen der Operator “+” zur Verfügung, mit dem zwei Listen aneinander gehängt werden können. Außerdem gibt es noch den unären Operator “#”, der für Mengen und Listen die Anzahl der Elemente

```

1  printSet := procedure(s) {
2      if (s == {}) {
3          return;
4      }
5      x := from(s);
6      print(x);
7      printSet(s);
8  };
9  s := { 13, 5, 7, 2, 4 };
10 printSet(s);

```

Abbildung 3.11: Menge elementweise ausdrucken.

berechnet. Schließlich kann man Elemente von Listen mit der Schreibweise

$$x := t[n];$$

indizieren. In diesem Fall muss t eine Liste sein, die mindestens die Länge n hat. Die obige Anweisung weist der Variablen x dann den Wert des n -ten Elementes der Liste t zu. Die obige Zuweisung läßt sich auch umdrehen: Mit

$$t[n] := x;$$

wird die Liste t so abgeändert, dass das n -te Element danach den Wert x hat. Im Gegensatz zu der Sprache C werden in SETLX Listen mit 1 beginnend indiziert, falls wir die beiden Befehle

```

1 := [ 1, 2, 3 ];
x := 1[1];

```

ausführen, hat x also anschließend den Wert 1.

```

1  a := [ 1, 2, 3 ];
2  b := [ 2, 3, 4, 5, 6 ];
3  c := { 5, 6, 7 };
4  // Aneinanderhängen von Tupeln mit +
5  print(a, " + ", b, " = ", a + b);
6  // Berechnung der Anzahl der Elemente einer Menge
7  print("# ", c, " = ", # c);
8  // Berechnung der Länge eines Tupels
9  print("# ", a, " = ", # a);
10 // Auswahl des dritten Elements von b
11 print(b, "[3] = ", b[3] );
12 // Überschreiben des 10. Elements von b
13 b[10] := 42;
14 print("b = ", b);
15 // Auswahl einer Teilliste
16 d := b(2..4);
17 print( "d = ", d);
18 x := 1; y := 2;
19 [ x, y ] := [ y, x ];
20 print("x = ", x, ", y = ", y);

```

Abbildung 3.12: Weitere Operatoren auf Tupeln und Mengen.

Das Programm `simple-tuple.stlx` in Abbildung 3.12 auf Seite 47 demonstriert die eben vorgestellten Operatoren. Zusätzlich sehen wir in Zeile 19, dass SETLX simultane Zuweisungen unterstützt. Das Programm produziert die folgende Ausgabe:

```
[1, 2, 3] + [2, 3, 4, 5, 6] = [1, 2, 3, 2, 3, 4, 5, 6]
# {5, 6, 7} = 3
# [1, 2, 3] = 3
[2, 3, 4, 5, 6][3] = 4
b = [2, 3, 4, 5, 6, om, om, om, om, 42]
d = [3, 4, 5]
x = 2, y = 1
```

3.5.1 Anwendung: *Sortieren durch Auswahl*

```

1  minSort := procedure(l) {
2      if (l == []) {
3          return [];
4      }
5      m := min(l);
6      return [ m ] + minSort( [ x in l | x != m ] );
7  };
8
9  l := [ 13, 5, 13, 7, 2, 4 ];
10 print( "sort( ", l, " ) = ", minSort(l) );
```

Abbildung 3.13: Implementierung des Algorithmus *Sortieren durch Auswahl*.

Als praktische Anwendung zeigen wir eine Implementierung des Algorithmus *Sortieren durch Auswahl*. Dieser Algorithmus, dessen Aufgabe es ist, eine gegebene Liste L zu sortieren, kann wie folgt beschrieben werden:

1. Falls L leer ist, so ist auch $\text{sort}(L)$ leer:

$$\text{sort}([]) = [].$$

2. Sonst berechnen wir das Minimum m von L :

$$m = \min(L).$$

Dann entfernen wir m aus der Liste L und sortieren die Restliste rekursiv:

$$\text{sort}(L) = [\min(L)] + \text{sort}([x \in L \mid x \neq \min(L)]).$$

Abbildung 3.13 auf Seite 48 zeigt das Programm `min-sort.stlx`, das diese Idee in SETLX umsetzt.

3.6 Kontroll-Strukturen

Die Sprache SETLX stellt alle Kontroll-Strukturen zur Verfügung, die in modernen Sprachen üblich sind. Wir haben “if”-Abfragen bereits mehrfach gesehen. In der allgemeinsten Form hat eine Fallunterscheidung die in Abbildung 3.14 auf Seite 49 gezeigte Struktur.

Hierbei steht test_i für einen Test, der “true” oder “false” liefert. Liefert der Test “true”, so wird der zugehörigen Anweisungen in body_i ausgeführt, andernfalls wird der nächste Test test_{i+1} versucht. Schlagen alle Tests fehl, so wird body_{n+1} ausgeführt.

Die Tests selber können dabei die binären Operatoren

```

1   if (test0) {
2       body0
3   } else if (test1) {
4       body1
5       ⋮
6   } else if (testn) {
7       bodyn
8   } else {
9       bodyn+1
10  }

```

Abbildung 3.14: Struktur der allgemeinen Fallunterscheidung.

“==”, “!=”, “>”, “<”, “>=”, “<=”, “in”,

verwenden. Dabei steht “==” für den Vergleich auf Gleichheit, “!=” für den Vergleich auf Ungleichheit. Für Zahlen führen die Operatoren “>”, “<”, “>=” und “<=” die selben Größenvergleiche durch wie in der Sprache *C*. Für Mengen überprüfen diese Operatoren analog, ob die entsprechenden Teilmengen-Beziehung erfüllt ist. Der Operator “in” überprüft, ob das erste Argument ein Element der als zweites Argument gegebene Menge ist: Der Test

$x \text{ in } S$

hat genau dann den Wert **true**, wenn $x \in S$ gilt. Aus den einfachen Tests, die mit Hilfe der oben vorgestellten Operatoren definiert werden können, können mit Hilfe der Junktoren “&&” (logisches *und*), “||” (logisches *oder*) und “not” (logisches *nicht*) komplexere Tests aufgebaut werden. Dabei bindet der Operator “||” am schwächsten und der Operator “!” bindet am stärksten. Ein Ausdruck der Form

`!a == b && b < c || x >= y`

wird also als

`((!(a == b)) && b < c) || x >= y`

geklammert. Damit haben die logischen Operatoren die selbe Präzedenz wie in der Sprache *C*.

SETLX unterstützt die Verwendung von Quantoren. Die Syntax für die Verwendung des Allquantors ist

`forall (x in s | cond)`

Hier ist s eine Menge und $cond$ ist ein Ausdruck, der von der Variablen x abhängt und einen Wahrheitswert als Ergebnis zurück liefert. Die Auswertung des oben angegebenen Allquantors liefert genau dann **true** wenn die Auswertung von $cond$ für alle Elemente der Menge s den Wert **true** ergibt. [Abbildung 3.15](#) auf Seite 50 zeigt das Programm `primes-forall.stlx`, welches die Primzahlen mit Hilfe eines Allquantors berechnet. Die Bedingung

`forall (x in divisors(p) | x in {1, p})`

trifft auf genau die Zahlen p zu, für die gilt, dass alle Teiler Elemente der Menge $\{1, p\}$ sind. Dies sind genau die Primzahlen.

Neben dem Allquantor gibt es noch den Existenzquantor. Die Syntax ist:

`exists (x in s | cond)`

Hierbei ist wieder eine s eine Menge und $cond$ ist ein Ausdruck, der zu **true** oder **false** ausgewertet werden kann. Falls es wenigstens ein x gibt, so dass die Auswertung von $cond$ **true** ergibt, liefert die Auswertung des Existenzquantors ebenfalls den Wert **true**.

```

1  isPrime := procedure(p) {
2      return forall (x in divisors(p) | x in {1, p});
3  };
4  divisors := procedure(p) {
5      return { t : t in {1..p} | p % t == 0 };
6  };
7  n := 100;
8  print([ p : p in [2..n] | isPrime(p) ]);

```

Abbildung 3.15: Berechnung der Primzahlen mit Hilfe eines Allquantors

Bemerkung: Liefert die Auswertung eines Ausdrucks der Form

`exists (x in s | cond)`

den Wert **true**, so wird **zusätzlich** der Variablen x ein Wert zugewiesen, für den die Bedingung *cond* erfüllt ist. Falls die Auswertung des Existenzquantors den Wert **false** ergibt, ändert sich der Wert von x nicht.

Switch-Blöcke

Als Alternative zur Fallunterscheidung mit Hilfe von **if-then-else**-Konstrukten gibt es noch den **switch**-Block. Ein solcher Block hat die in Abbildung 3.16 auf Seite 50 gezeigte Struktur. Bei der Abarbeitung werden der Reihe nach die Tests $test_1, \dots, test_n$ ausgewertet. Für den ersten Test $test_i$, dessen Auswertung den Wert **true** ergibt, wird der zugehörige Block $body_i$ ausgeführt. Nur dann, wenn alle Tests $test_1, \dots, test_n$ scheitern, wird der Block $body_{n+1}$ hinter dem Schlüsselwort **default** ausgeführt. Den selben Effekt könnte man natürlich auch mit einer **if-else if-...-else if-else**-Konstruktion erreichen, nur ist die Verwendung eines **switch**-Blocks oft übersichtlicher.

```

1  switch {
2      case test1 : body1
3      :
4      case testn : bodyn
5      default: bodyn+1
6  }

```

Abbildung 3.16: Struktur eines **switch**-Blocks

Abbildung 3.17 zeigt das Programm **switch.stlx**, bei dem es darum geht, in Abhängigkeit von der letzte Ziffer einer Zahl eine Meldung auszugeben. Bei der Behandlung der Aussagenlogik werden wir noch realistischere Anwendungs-Beispiele für den **switch**-Block kennenlernen.

In der Sprache **C** gibt es eine analoge Konstruktion. In **C** ist es so, dass nach einem Block, der nicht durch einen **break**-Befehl abgeschlossen wird, auch alle folgenden Blocks ausgeführt werden. Dies ist in SETLX anders: Dort wird immer genau ein Block ausgeführt.

Neben dem **switch**-Block gibt es noch dem **match**-Block, den wir allerdings erst später diskutieren werden.

3.6.1 Schleifen

Es gibt in SETLX Kopf-gesteuerte Schleifen (**while**-Schleifen) und Schleifen, die über die Elemente einer Menge iterieren (**for**-Schleifen). Wir diskutieren diese Schleifenformen jetzt im Detail.

```

1  print("Zahl eingeben:");
2  n := read();
3  m := n % 10;
4  switch {
5      case m == 0 : print("letzte Ziffer ist 0");
6      case m == 1 : print("letzte Ziffer ist 1");
7      case m == 2 : print("letzte Ziffer ist 2");
8      case m == 3 : print("letzte Ziffer ist 3");
9      case m == 4 : print("letzte Ziffer ist 4");
10     case m == 5 : print("letzte Ziffer ist 5");
11     case m == 6 : print("letzte Ziffer ist 6");
12     case m == 7 : print("letzte Ziffer ist 7");
13     case m == 8 : print("letzte Ziffer ist 8");
14     case m == 9 : print("letzte Ziffer ist 9");
15     default      : print("impossible");
16 }

```

Abbildung 3.17: Anwendung eines `switch`-Blocks

while-Schleifen

Die allgemeine Syntax der `while`-Schleife ist in Abbildung 3.18 auf Seite 51 gezeigt. Hierbei ist *test* ein Ausdruck, der zu Beginn ausgewertet wird und der “true” oder “false” ergeben muss. Ergibt die Auswertung “false”, so ist die Auswertung der `while`-Schleife bereits beendet. Ergibt die Auswertung allerdings “true”, so wird anschließend *body* ausgewertet. Danach beginnt die Auswertung der Schleife dann wieder von vorne, d.h. es wird wieder *test* ausgewertet und danach wird abhängig von dem Ergebnis dieser wieder *body* ausgewertet. Das ganze passiert so lange, bis irgendwann einmal die Auswertung von *test* den Wert “false” ergibt. Die von SETLX unterstützten `while`-Schleifen funktionieren genauso wie in der Sprache C.

```

while (test) {
    body
}

```

Abbildung 3.18: Struktur der `while`-Schleife

Abbildung 3.19 auf Seite 51 zeigt das Programm `primes-while.stlx`, das die Primzahlen mit Hilfe einer `while`-Schleife berechnet. Hier ist die Idee, dass eine Zahl genau dann Primzahl ist, wenn es keine kleinere Primzahl gibt, die diese Zahl teilt.

```

1  n := 100;
2  primes := {};
3  p := 2;
4  while (p <= n) {
5      if (forall (t in primes | p % t != 0)) {
6          print(p);
7          primes := primes + { p };
8      }
9      p := p + 1;
10 }

```

Abbildung 3.19: Iterative Berechnung der Primzahlen.

for-Schleifen

Die allgemeine Syntax der `for`-Schleife ist in Abbildung 3.20 auf Seite 52 gezeigt. Hierbei ist s eine Menge (oder eine Liste) und x der Name einer Variablen. Diese Variable wird nacheinander mit allen Werten aus der Menge s belegt und anschließend wird mit dem jeweiligen Wert von x der Schleifenrumpf *body* ausgeführt.

```
for (x in s) {  
    body  
}
```

Abbildung 3.20: Struktur der `for`-Schleife.

Abbildung 3.21 auf Seite 52 zeigt das Programm `primes-for.stlx`, das die Primzahlen mit Hilfe einer `for`-Schleife berechnet. Der dabei verwendete Algorithmus ist als das *Sieb des Eratosthenes* bekannt. Das funktioniert wie folgt: Sollen alle Primzahlen kleiner oder gleich n berechnet werden, so wird zunächst ein Tupel der Länge n gebildet, dessen i -tes Element den Wert i hat. Das passiert in Zeile 3. Anschließend werden alle Zahlen, die Vielfache von 2, 3, 4, \dots sind, aus der Menge der Primzahlen entfernt, indem die Zahl, die an dem entsprechenden Index in der Liste `primes` steht, auf 0 gesetzt wird. Dazu sind zwei Schleifen erforderlich: Die äußere `for`-Schleife iteriert i über alle Werte von 2 bis n . Die innere `while`-Schleife iteriert dann für gegebenes i über alle Werte $i \cdot j$, für die das Produkt $i \cdot j \leq n$ ist. Schließlich werden in der letzten `for`-Schleife in den Zeilen 14 bis 18 alle die Indizes i ausgedruckt, für die `primes(i)` nicht auf 0 gesetzt worden ist, denn das sind genau die Primzahlen.

```
1  n := 100;  
2  primes := [1 .. n];  
3  for (i in [2 .. n]) {  
4      j := 2;  
5      while (i * j <= n) {  
6          primes[i * j] := 0;  
7          j := j + 1;  
8      }  
9  }  
10 print({ i : i in [2 .. n] | primes[i] > 0 });
```

Abbildung 3.21: Berechnung der Primzahlen nach Eratosthenes.

Der Algorithmus aus Abbildung 3.21 kann durch die folgenden Beobachtungen noch verbessert werden:

1. Es reicht aus, wenn j mit i initialisiert wird, denn alle kleineren Vielfachen von i wurden bereits vorher auf 0 gesetzt.
2. Falls in der äußeren Schleife die Zahl i keine Primzahl ist, so bringt es nichts mehr, die innere `while`-Schleife in den Zeilen 9 bis 11 zu durchlaufen, denn alle Indizes, für die dort `primes[i*j]` auf 0 gesetzt wird, sind schon bei dem vorherigen Durchlauf der äußeren Schleife, bei der `primes[i]` auf 0 gesetzt wurde, zu 0 gesetzt worden.

Abbildung 3.22 auf Seite 53 zeigt das Programm `primes-eratosthenes.stlx`, das diese Ideen umsetzt. Um den Durchlauf der inneren `while`-Schleife in dem Fall, dass `primes[i] = 0` ist, zu überspringen, haben wir den Befehl `“continue”` benutzt. Der Aufruf von `“continue”` bricht die Abarbeitung des Schleifen-Rumpfs für den aktuellen Wert von i ab, weist der Variablen i den nächsten Wert aus `[1..n]` zu und fährt dann mit der Abarbeitung der Schleife in Zeile 4 fort. Der Befehl `“continue”` verhält sich also genauso, wie der Befehl `“continue”` in der Sprache *C*.

```

1  n := 10000;
2  primes := [1 .. n];
3  for (i in [2 .. n/2]) {
4      if (primes[i] == 0) {
5          continue;
6      }
7      j := i;
8      while (i * j <= n) {
9          primes[i * j] := 0;
10         j := j + 1;
11     }
12 }
13 print({ i : i in [2 .. n] | primes[i] > 0 });

```

Abbildung 3.22: Effizientere Berechnung der Primzahlen nach Eratosthenes.

3.6.2 Fixpunkt-Algorithmen

Angenommen, wir wollen in der Menge \mathbb{R} der reellen Zahlen die Gleichung

$$x = \cos(x)$$

lösen. Ein naives Verfahren, das hier zum Ziel führt, basiert auf der Beobachtung, dass die Folge $(x_n)_n$, die durch

$$x_0 := 0 \text{ und } x_{n+1} := \cos(x_n) \text{ für alle } n \in \mathbb{N}$$

definiert ist, gegen eine Lösung der obigen Gleichung konvergiert. Damit führt das in Abbildung 3.23 auf Seite 53 angegebene Programm `solve.stlx` zum Ziel.

```

1  x := 0.0;
2  while (true) {
3      old_x := x;
4      x := cos(x);
5      print(x);
6      if (abs(x - old_x) < 1.0e-13) {
7          print("x = ", x);
8          break;
9      }
10 }

```

Abbildung 3.23: Lösung der Gleichung $x = \cos(x)$ durch Iteration.

Bei dieser Implementierung wird die Schleife in dem Moment abgebrochen, wenn die Werte von `x` und `old_x` nahe genug beieinander liegen. Dieser Test kann aber am Anfang der Schleife noch gar nicht durchgeführt werden, weil da die Variable `old_x` noch gar keinen Wert hat. Daher brauchen wir hier das Kommando “`break`”. Dieses bricht die Schleife ab. In der Sprache *C* hat das Kommando “`break`” die selbe Funktion.

Abbildung 3.24 zeigt das Programm `fixpoint.stlx`. Hier ist eine generische Funktion `solve` implementiert, die zwei Argumente verarbeitet:

1. Das Argument `f` ist die Funktion, für die ein Fixpunkt berechnet werden soll.
2. Das Argument `x0` gibt den Startwert der Fixpunkt-Iteration an.

```

1  solve := procedure(f, x0) {
2      x := x0;
3      for (n in [1 .. 10000]) {
4          oldX := x;
5          x := f(x);
6          if (abs(x - oldX) < 1.0e-15) {
7              return x;
8          }
9      }
10 }
11 print("solution to x = cos(x): ", solve(cos, 0));
12 print("solution to x = 1/(1+x): ", solve(x |-> 1.0/(1+x), 0));

```

Abbildung 3.24: Eine generische Implementierung des Fixpunkt-Algorithmus

In Zeile 11 haben wir die Funktion `solve` aufgerufen, um die Gleichung $x = \cos(x)$ zu lösen. In Zeile 12 lösen wir die Gleichung

$$x = \frac{1}{1+x}$$

die zu der quadratischen Gleichung $x^2 + x = 1$ äquivalent ist. Beachten Sie, dass wir die Funktion $x \mapsto \frac{1}{1+x}$ in SETLX durch den Ausdruck

`x |-> 1.0/(1+x)`

definieren können. Hier habe ich die Fließkomma-Zahl 1.0 verwendet, das SETLX sonst mit rationalen Zahlen rechnen würde, was wesentlich aufwendiger ist.

Ganz nebenbei zeigt das obige Beispiel auch, dass Sie in SETLX nicht nur mit ganzen, sondern auch mit reellen Zahlen rechnen können. Eine Zahlen-Konstante, die den Punkt “.” enthält, wird automatisch als reelle Zahl erkannt und auch so abgespeichert. In SETLX stehen unter anderem die folgenden reellen Funktionen zur Verfügung:

1. Der Ausdruck `sin(x)` berechnet den Sinus von x . Außerdem stehen die trigonometrischen Funktionen `cos(x)` und `tan(x)` zur Verfügung. Die Umkehr-Funktionen der trigonometrischen Funktionen sind `asin(x)`, `acos(x)` und `atan(x)`.

Der Sinus Hyperbolicus von x wird durch `sinh(x)` berechnet. Entsprechend berechnet `cosh(x)` den Kosinus Hyperbolicus und `tanh(x)` den Tangens Hyperbolicus.

2. Der Ausdruck `exp(x)` berechnet die Potenz zur Basis e , es gilt

$$\exp(x) = e^x.$$

3. Der Ausdruck `log(x)` berechnet den natürlichen Logarithmus von x . Der Logarithmus zur Basis 10 von x wird durch `log10(x)` berechnet.

4. Der Ausdruck `abs(x)` berechnet den Absolut-Betrag von x , während `signum(x)` das Vorzeichen von x berechnet.

5. Der Ausdruck `sqrt(x)` berechnet die Quadrat-Wurzel von x , es gilt

$$\text{sqrt}(x) = \sqrt{x}.$$

6. Der Ausdruck `cbirt(x)` berechnet die dritte Wurzel von x , es gilt

$$\text{cbirt}(x) = \sqrt[3]{x}.$$

7. Der Ausdruck `ceil(x)` berechnet die kleinste ganze Zahl, die größer oder gleich x ist, es gilt

$$\text{ceil}(x) = \min(\{z \in \mathbb{Z} \mid z \geq x\}).$$

8. Der Ausdruck `floor(x)` berechnet die größte ganze Zahl, die kleiner oder gleich x ist, es gilt

$$\text{floor}(x) = \max(\{z \in \mathbb{Z} \mid z \leq x\}).$$

9. Der Ausdruck `round(x)` rundet x zu einer ganzen Zahl.

Darüber hinaus unterstützt SETLX die Verwendung von rationalen Zahlen. Die Eingabe von

```
1/2 + 1/3;
```

liefert das Ergebnis $5/6$.

3.6.3 Verschiedenes

Der Interpreter bietet die Möglichkeit, komplexe Programme zu laden. Der Befehl

```
load(file);
```

lädt das Programm, das sich in der Datei *file* befindet und führt die in dem Programm vorhandenen Befehle aus. Führen wir beispielsweise den Befehl

```
load("primes-forall.stlx");
```

im Interpreter aus und enthält die Datei `"primes-forall.stlx"` das in Abbildung 3.15 auf Seite 50 gezeigte Programm, so können wir anschließend mit den in dieser Datei definierten Variablen arbeiten. Beispielsweise liefert der Befehl

```
print(isPrime);
```

die Ausgabe:

```
procedure (p) { return forall (x in divisors(p) | x in {1, p}); }
```

Zeichenketten, auch bekannt als *Strings*, werden in SETLX in doppelte Hochkommata gesetzt. Der Operator `"+"` kann dazu benutzt werden, zwei Strings aneinander zu hängen, der Ausdruck

```
"abc" + "uvw";
```

liefert also das Ergebnis

```
"abcuvw".
```

Zusätzlich kann eine natürliche Zahl n mit einem String s über den Multiplikations-Operator `"*"` verknüpft werden. Der Ausdruck

```
n * s;
```

liefert als Ergebnis die n -malige Verkettung von s . Beispielsweise ist das Ergebnis von

```
3 * "abc";
```

der String

```
"abcabcabc".
```

3.7 Fallstudie: Berechnung von Wahrscheinlichkeiten

Wir wollen in diesem kurzen Abschnitt zeigen, wie sich Wahrscheinlichkeiten für die Poker-Variante *Texas Hold'em* berechnen lassen. Bei dieser Poker-Variante gibt es insgesamt 52 Karten. Jeder dieser Karten hat eine Farbe, die ein Element der Menge

$$\text{suits} = \{\clubsuit, \spadesuit, \heartsuit, \diamondsuit\}$$

ist und außerdem einen Wert hat, der ein Element der Menge

$$\text{values} = \{2, 3, 4, 5, 6, 7, 8, 9, 10, \text{Jack}, \text{Queen}, \text{King}, \text{Ace}\}$$

ist. Die einzelnen Karten können dann mathematisch als Paare dargestellt werden. Die gesamte Menge von Karten wäre dann beispielsweise wie folgt gegeben:

$$\text{deck} = \{\langle v, s \rangle \mid v \in \text{values} \wedge s \in \text{suits}\}.$$

Jeder Spieler bekommt zunächst zwei Karten. Diese beiden Karten werden als *Preflop* oder *Hole* bezeichnet. Nach einer Bietphase werden anschließend drei weitere Karten, der sogenannte *Flop*, auf den Tisch gelegt. Wir nehmen an, dass ein Spieler zunächst die Karten $\{\langle 3, \clubsuit \rangle, \langle 3, \spadesuit \rangle\}$ bekommen hat und nun wissen möchte, wie groß die Wahrscheinlichkeit dafür ist, dass beim Flop eine weitere 3 auf den Tisch gelegt wird. Um diese Wahrscheinlichkeit zu berechnen, muss der Spieler die Anzahl aller Flops, bei denen eine weitere 3 erscheinen kann, durch die gesamte Anzahl der Flops teilen. Das in Abbildung 3.25 gezeigte Programm führt diese Berechnung durch.

```

1 suits := { "c", "h", "d", "s" };
2 values := { "2", "3", "4", "5", "6", "7", "8", "9",
3           "T", "J", "Q", "K", "A" };
4 deck := { [ v, s ] : v in values, s in suits };
5 hole := { [ "3", "c" ], [ "3", "s" ] };
6 rest := deck - hole;
7 flops := { { k1, k2, k3 } : k1 in rest, k2 in rest, k3 in rest
8                       | #{ k1, k2, k3 } == 3 };
9 print(#flops);
10 trips := { f : f in flops | [ "3", "d" ] in f || [ "3", "h" ] in f };
11 print(1.0 * #trips / #flops);

```

Abbildung 3.25: Berechnung von Wahrscheinlichkeiten im Poker

1. In Zeile 1 definieren wir die Menge **suits** der Farben einer Karte.
2. In Zeile 2 und 3 definieren wir die Menge **values** der möglichen Werte einer Karte.
3. In Zeile 4 stellen wir die Menge aller Karten, die wir mit *deck* bezeichnen, als Menge von Paaren dar, wobei die erste Komponente der Paare den Wert und die zweite Komponente die Farbe der Karte angibt.
4. Die in Zeile 5 definierte Menge *hole* stellt die Menge der Karten des Spielers dar.
5. Die verbleibenden Karten bezeichnen wir in Zeile 6 als *rest*.
6. In Zeile 7 und 8 berechnen wir die Menge aller möglichen Flops. Da die Reihenfolge der Karten im Flop keine Rolle spielt, verwenden wir zur Darstellung der einzelnen Flops eine Menge. Dabei müssen wir aber darauf achten, dass der Flop auch wirklich aus drei verschiedenen Karten besteht. Daher haben wir bei der Auswahl der Menge die Bedingung

$$\# \{ k1, k2, k3 \} == 3$$

zu beachten.

7. Die Teilmenge der Flops, in denen mindestens eine weitere 3 auftritt, wird in Zeile 10 berechnet.
8. Schließlich berechnet sich die Wahrscheinlichkeit für eine 3 im Flop als das Verhältnis der Anzahl der günstigen Fälle zu der Anzahl der möglichen Fälle. Wir müssen also nur die Anzahl der Elemente der entsprechenden Mengen ins Verhältnis setzen.

Hier gibt es aber noch eine Klippe, die umschifft werden muss: Da die Variablen `#Trips` und `#Flops` mit ganzen Zahlen belegt sind, würde die Division `#Trips / #Flops` als ganzzahlige Division ausgeführt und das Ergebnis wäre eine rationale Zahl. Daher ist es erforderlich, zunächst eine der beiden Zahlen in eine Fließkommazahl zu konvertieren, damit der Quotient als Fließkommazahl berechnet wird. Dies können wir beispielsweise dadurch erreichen, dass wir den Dividenten `#Trips` mit der Fließkommazahl 1.0 multiplizieren.

Lassen wir das Programm laufen, so sehen wir, dass die Wahrscheinlichkeit, ein Pocket-Paar im Flop auf Trips zu verbessern, bei etwa 11,8% liegt.

Bemerkung: Die Berechnung von Wahrscheinlichkeiten ist in der oben dargestellten Weise nur dann möglich, wenn die im Laufe der Berechnung auftretenden Mengen so klein sind, dass sie sich explizit im Rechner darstellen lassen. Wenn diese Voraussetzung nicht mehr erfüllt ist, können die gesuchten Wahrscheinlichkeiten mit Hilfe der im zweiten Semester vorgestellten *Monte-Carlo-Methode* berechnet werden.

3.8 Fallstudie: Berechnung von Pfaden

Wir wollen dieses Kapitel mit einer praktisch relevanten Anwendung der Sprache SETLX abschließen. Dazu betrachten wir das Problem, Pfade in einem *Graphen* zu bestimmen. Abstrakt gesehen beinhaltet ein Graph die Information, zwischen welchen Punkten es direkte Verbindungen gibt. Zur Vereinfachung wollen wir zunächst annehmen, dass die einzelnen Punkte durch Zahlen identifiziert werden. Dann können wir eine direkte Verbindung zwischen zwei Punkten durch ein Paar von Zahlen darstellen. Den Graphen selber stellen wir als eine Menge solcher Paaren dar. Wir betrachten ein Beispiel. Sei R wie folgt definiert:

$$R := \{ \langle 1, 2 \rangle, \langle 2, 3 \rangle, \langle 1, 3 \rangle, \langle 2, 4 \rangle, \langle 4, 5 \rangle \}.$$

In diesem Graphen haben wir die Punkte 1, 2, 3, 4 und 5. Eine Darstellung dieses Graphen finden Sie in Abbildung 3.26. Beachten Sie, dass die Verbindungen in diesem Graphen *Einbahn-Straßen* sind: Wir haben zwar eine Verbindung von 1 nach 2, aber keine Verbindung von 2 nach 1.

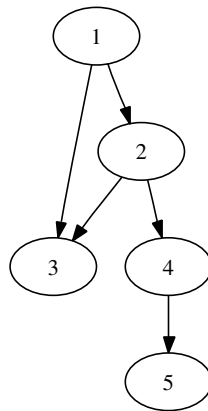


Abbildung 3.26: Ein einfacher Graph.

In dem Graphen sind nur die unmittelbaren Verbindungen zwischen zwei Punkten verzeichnet. Es gibt aber unter Umständen auch noch andere Verbindungen. Beispielsweise gibt es eine unmittelbare Verbindung von 1 nach 3. Es gibt darüber hinaus noch einen Pfad von 1 nach 3, der über den Punkt 2 geht. Unser Ziel in diesem Abschnitt ist es einen Algorithmus zu entwickeln, der überprüft, ob zwischen zwei Punkten eine Verbindung existiert und gegebenenfalls berechnet. Dazu entwickeln wir zunächst einen Algorithmus, der nur überprüft, ob es eine Verbindung zwischen zwei Punkten gibt und erweitern diesen Algorithmus dann später so, dass diese Verbindung auch berechnet wird.

3.8.1 Berechnung des transitiven Abschlusses einer Relation

Als erstes bemerken wir, dass ein Graph R nichts anderes ist als eine binäre Relation. Um feststellen zu können, ob es zwischen zwei Punkten eine Verbindung gibt, müssen wir den transitiven Abschluß R^+ der Relation R bilden. Wir haben bereits früher in dem Kapitel zur Mengenlehre gezeigt, dass R^+ wie folgt berechnet werden kann:

$$R^+ = \bigcup_{i=1}^{\infty} R^i = R \cup R^2 \cup R^3 \cup \dots$$

Auf den ersten Blick betrachtet sieht diese Formel so aus, als ob wir unendlich lange rechnen müßten. Aber versuchen wir einmal, diese Formel anschaulich zu verstehen. Zunächst steht da R . Das sind die Verbindungen, die unmittelbar gegeben sind. Als nächstes steht dort R^2 und das ist $R \circ R$. Es gilt aber

$$R \circ R = \{ \langle x, z \rangle \mid \exists y: \langle x, y \rangle \in R \wedge \langle y, z \rangle \in R \}$$

In R^2 sind also alle die Pfade enthalten, die aus zwei direkten Verbindungen zusammengesetzt sind. Allgemein läßt sich durch Induktion sehen, dass R^n alle die Pfade enthält, die aus n direkten Verbindungen zusammengesetzt sind. Nun ist die Zahl der Punkte, die wir haben, endlich. Sagen wir mal, dass es k Punkte sind. Dann macht es keinen Sinn solche Pfade zu betrachten, die aus mehr als $k-1$ direkten Verbindungen zusammengesetzt sind, denn wir wollen ja nicht im Kreis herum laufen. Damit kann dann aber die Formel zur Berechnung des transitiven Abschlusses vereinfacht werden:

$$R^+ = \bigcup_{i=1}^{k-1} R^i.$$

Diese Formel könnten wir tatsächlich so benutzen. Es ist aber noch effizienter, einen Fixpunkt-Algorithmus zu verwenden. Dazu zeigen wir zunächst, dass der transitive Abschluß R^+ die folgende Fixpunkt-Gleichung erfüllt:

$$R^+ = R \cup R \circ R^+. \quad (3.1)$$

Wir erinnern hier daran, dass wir vereinbart haben, dass der Operator \circ stärker bindet als der Operator \cup , so dass der Ausdruck $R \cup R \circ R^+$ als $R \cup (R \circ R^+)$ zu lesen ist. Die Fixpunkt-Gleichung 3.1 läßt sich algebraisch beweisen. Es gilt

$$\begin{aligned} & R \cup R \circ R^+ \\ = & R \cup R \circ \bigcup_{i=1}^{\infty} R^i \\ = & R \cup R \circ (R^1 \cup R^2 \cup R^3 \cup \dots) \\ = & R \cup (R \circ R^1 \cup R \circ R^2 \cup R \circ R^3 \cup \dots) \quad \text{Distributiv-Gesetz} \\ = & R \cup (R^2 \cup R^3 \cup R^4 \cup \dots) \quad \text{Potenz-Gesetz} \\ = & R^1 \cup (R^2 \cup R^3 \cup R^4 \cup \dots) \\ = & \bigcup_{i=1}^{\infty} R^i \\ = & R^+ \end{aligned}$$

Die Gleichung 3.1 kann benutzt werden um den transitiven Abschluß iterativ zu berechnen. Wir definieren eine Folge $(T_n)_{n \in \mathbb{N}}$ durch Induktion folgt:

$$\begin{aligned} \text{I.A. } n=1: & \quad T_1 := R \\ \text{I.S. } n \mapsto n+1: & \quad T_{n+1} := R \cup R \circ T_n. \end{aligned}$$

Die Relationen T_n lassen sich auf die Relation R zurückführen:

1. $T_1 = R$.
2. $T_2 = R \cup R \circ T_1 = R \cup R \circ R = R^1 \cup R^2$.

$$\begin{aligned}
3. \quad T_3 &= R \cup R \circ T_2 \\
&= R \cup R \circ (R^1 \cup R^2) \\
&= R^1 \cup R^2 \cup R^3.
\end{aligned}$$

Allgemein können wir durch vollständige Induktion über $n \in \mathbb{N}$ beweisen, dass

$$T_n = \bigcup_{i=1}^n R^i$$

gilt. Der Induktions-Anfang folgt unmittelbar aus der Definition von T_1 . Um den Induktions-Schritt durchzuführen, betrachten wir

$$\begin{aligned}
T_{n+1} &= R \cup R \circ T_n && \text{gilt nach Definition} \\
&= R \cup R \circ \left(\bigcup_{i=1}^n R^i \right) && \text{gilt nach Induktions-Voraussetzung} \\
&= R \cup R^2 \cup \dots \cup R^{n+1} && \text{Distributiv-Gesetz} \\
&= R^1 \cup \dots \cup R^{n+1} \\
&= \bigcup_{i=1}^{n+1} R^i && \square
\end{aligned}$$

Die Folge $(T_n)_{n \in \mathbb{N}}$ hat eine weitere nützliche Eigenschaft: Sie ist *monoton steigend*. Allgemein nennen wir eine Folge von Mengen $(X_n)_{n \in \mathbb{N}}$ *monoton steigend*, wenn

$$\forall n \in \mathbb{N} : X_n \subseteq X_{n+1}$$

gilt, wenn also die Mengen X_n mit wachsendem Index n immer größer werden. Die Monotonie der Folge $(T_n)_{n \in \mathbb{N}}$ folgt aus der gerade bewiesenen Eigenschaft $T_n = \bigcup_{i=1}^n R^i$, denn es gilt

$$\begin{aligned}
T_n &\subseteq T_{n+1} \\
\Leftrightarrow \bigcup_{i=1}^n R^i &\subseteq \bigcup_{i=1}^{n+1} R^i \\
\Leftrightarrow \bigcup_{i=1}^n R^i &\subseteq \bigcup_{i=1}^n R^i \cup R^{n+1}
\end{aligned}$$

und die letzte Formel ist offenbar wahr. Ist nun die Relation R endlich, so ist natürlich auch R^+ eine endliche Menge. Da die Folge T_n aber in dieser Menge liegt, denn es gilt ja

$$T_n = \bigcup_{i=1}^n R^i \subseteq \bigcup_{i=1}^{\infty} R^i = R^+ \quad \text{für alle } n \in \mathbb{N},$$

können die Mengen T_n nicht beliebig groß werden. Aufgrund der Monotonie der Folge $(T_n)_{n \in \mathbb{N}}$ muss es daher einen Index k geben, ab dem die Mengen T_n alle gleich sind:

$$\forall n \in \mathbb{N} : (n \geq k \rightarrow T_n = T_k).$$

Berücksichtigen wir die Gleichung $T_n = \bigcup_{i=1}^n R^i$, so haben wir

$$T_n = \bigcup_{i=1}^n R^i = \bigcup_{i=1}^k R^i = T_k \quad \text{für alle } n \geq k.$$

Daraus folgt dann aber, dass

$$T_n = \bigcup_{i=1}^n R^i = \bigcup_{i=1}^{\infty} R^i = R^+ \quad \text{für alle } n \geq k$$

gilt. Der Algorithmus zur Berechnung von R^+ sieht nun so aus, dass wir die Iteration

$$T_{n+1} := R \cup R \circ T_n$$

solange durchführen bis $T_{n+1} = T_n$ gilt, denn dann gilt auch $T_n = R^+$.

```

1  closure := procedure(r) {
2      t := r;
3      while (true) {
4          oldT := t;
5          t := r + product(r, t);
6          if (t == oldT) {
7              return t;
8          }
9      }
10 };
11 product := procedure(r1, r2) {
12     return { [x,z] : [x,y] in r1, [y,z] in r2 };
13 };
14 r := { [1,2], [2,3], [1,3], [2,4], [4,5] };
15 print( "r = ", r );
16 print( "computing transitive closure of r" );
17 t := closure(r);
18 print( "r+ = ", t );

```

Abbildung 3.27: Berechnung des transitiven Abschlusses.

Das Programm `transitive-closure.stlx` in Abbildung 3.27 auf Seite 60 zeigt eine Implementierung dieses Gedankens. Lassen wir dieses Programm laufen, so erhalten wir als Ausgabe:

```

R = {[2, 3], [4, 5], [1, 3], [2, 4], [1, 2]}
R+ = {[1, 5], [2, 3], [4, 5], [1, 4], [1, 3], [2, 4], [1, 2], [2, 5]}

```

Der transitive Abschluß R^+ der Relation R läßt sich jetzt anschaulich interpretieren: Er enthält alle Paare $\langle x, y \rangle$, für die es einen *Pfad* von x nach y gibt. Ein Pfad von x nach y ist dabei eine Liste der Form

$$[x_1, x_2, \dots, x_n],$$

für die $x = x_1$ und $y = x_n$ gilt und für die außerdem

$$\langle x_i, x_{i+1} \rangle \in R \quad \text{für alle } i = 1, \dots, n-1 \text{ gilt.}$$

Die Funktion $product(r_1, r_2)$ berechnet das relationale Produkt $r_1 \circ r_2$ nach der Formel

$$r_1 \circ r_2 = \{ \langle x, z \rangle \mid \exists y : \langle x, y \rangle \in r_1 \wedge \langle y, z \rangle \in r_2 \}.$$

Die Implementierung dieser Prozedur zeigt die allgemeine Form der Mengen-Definition durch Iteratoren in SET-LX. Allgemein können wir eine Menge durch den Ausdruck

$$\{ \text{expr} : [x_1^{(1)}, \dots, x_{n(1)}^{(1)}] \text{ in } s_1, \dots, [x_1^{(k)}, \dots, x_{n(k)}^{(k)}] \text{ in } s_k \mid \text{cond} \}$$

definieren. Dabei muss s_i für alle $i = 1, \dots, k$ eine Menge von Listen der Länge $n(i)$ sein. Bei der Auswertung dieses Ausdrucks werden für die Variablen $x_1^{(i)}, \dots, x_{n(i)}^{(i)}$ die Werte eingesetzt, die die entsprechenden Komponenten der Listen haben, die in der Menge s_i auftreten. Beispielsweise würde die Auswertung von

```

s1 := { [ 1, 2, 3 ], [ 5, 6, 7 ] };
s2 := { [ "a", "b" ], [ "c", "d" ] };
m := { [ x1, x2, x3, y1, y2 ] : [ x1, x2, x3 ] in s1, [ y1, y2 ] in s2 };

```

für M die Menge

```

{ [1, 2, 3, "a", "b"], [5, 6, 7, "c", "d"],
  [1, 2, 3, "c", "d"], [5, 6, 7, "a", "b"] }

```

berechnen.

3.8.2 Berechnung der Pfade

Als nächstes wollen wir das Programm zur Berechnung des transitiven Abschlusses so erweitern, dass wir nicht nur feststellen können, dass es einen Pfad zwischen zwei Punkten gibt, sondern dass wir diesen auch berechnen können. Die Idee ist, dass wir statt des relationalen Produkts, das für zwei Relationen definiert ist, ein sogenanntes *Pfad-Produkt*, das auf Mengen von Pfaden definiert ist, berechnen. Vorab führen wir für Pfade, die wir ja durch Listen repräsentieren, drei Begriffe ein.

1. Die Funktion $first(p)$ liefert den ersten Punkt der Liste p :

$$first([x_1, \dots, x_m]) = x_1.$$

2. Die Funktion $last(p)$ liefert den letzten Punkt der Liste p :

$$last([x_1, \dots, x_m]) = x_m.$$

3. Sind $p = [x_1, \dots, x_m]$ und $q = [y_1, \dots, y_n]$ zwei Pfade mit $first(q) = last(p)$, dann definieren wir die Summe von p und q als

$$p \oplus q := [x_1, \dots, x_m, y_2, \dots, y_n].$$

Sind nun P_1 und P_2 Mengen von Pfaden, so definieren wir das *Pfad-Produkt* von P_1 und P_2 als

$$P_1 \bullet P_2 := \{ p_1 \oplus p_2 \mid p_1 \in P_1 \wedge p_2 \in P_2 \wedge last(p_1) = first(p_2) \}.$$

```

1  closure := procedure(r) {
2      p := r;
3      while (true) {
4          oldP := p;
5          p := r + pathProduct(r, p);
6          if (p == oldP) {
7              return p;
8          }
9      }
10 };
11 pathProduct := procedure(p, q) {
12     return { add(x, y) : x in p, y in q | x[#x] == y[1] };
13 };
14 add := procedure(p, q) {
15     return p + q[2..];
16 };
17 r := { [1,2], [2,3], [1,3], [2,4], [4,5] };
18 print( "r = ", r );
19 print( "computing all paths" );
20 p := closure(r);
21 print( "p = ", p );

```

Abbildung 3.28: Berechnung aller Verbindungen.

Damit können wir das Programm in Abbildung 3.27 so abändern, dass alle möglichen Verbindungen zwischen zwei Punkten berechnet werden. Abbildung 3.28 zeigt das resultierende Programm `path.stlx`. Leider funktioniert das Programm dann nicht mehr, wenn der Graph Zyklen enthält. Abbildung 3.29 zeigt einen Graphen, der einen Zyklus enthält. In diesem Graphen gibt es unendlich viele Pfade, die von dem Punkt 1 zu dem Punkt

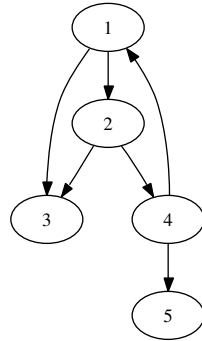


Abbildung 3.29: Ein zyklischer Graph.

2 führen:

$[1, 2], [1, 2, 4, 1, 2], [1, 2, 4, 1, 2, 4, 1, 2], [1, 2, 4, 1, 2, 4, 1, 2, 4, 1, 4], \dots$

Offenbar sind die Pfade unwichtig, die einen Punkt mehrfach enthalten und die daher zyklisch sind. Solche Pfade sollten wir bei der Berechnung des Pfad-Produktes eliminieren.

```

1  pathProduct := procedure(p, q) {
2      return { add(x,y) : x in p, y in q | x[#x] == y[1] && noCycle(x,y) };
3  };
4  noCycle := procedure(l1, l2) {
5      return #({ x : x in l1 } * { x : x in l2 }) == 1;
6  };

```

Abbildung 3.30: Berechnung aller Verbindungen in zyklischen Graphen

Abbildung 3.30 zeigt einen Ausschnitt des geänderten Programms `path-cyclic.stlx`, das auch für zyklische Graphen funktioniert.

1. In Zeile 2 berücksichtigen wir nur die Pfade $x \oplus y$, die nicht zyklisch sind.
2. In Zeile 5 überprüfen wir, ob die Konkatenation $l_1 \oplus l_2$ zyklisch ist. Die Kombination von l_1 und l_2 ist genau dann zyklisch, wenn die Listen l_1 und l_2 mehr als ein gemeinsames Element enthalten. Die Listen l_1 und l_2 enthalten mindestens ein gemeinsames Element, denn wir verknüpfen diese beiden Listen ja nur dann, wenn das letzte Element der Liste l_1 mit dem ersten Element der Liste l_2 übereinstimmt. Wenn es nun noch einen weiteren Punkt geben würde, der sowohl in l_1 als auch in l_2 auftreten würde, dann wäre der Pfad $l_1 \oplus l_2$ zyklisch.

In den meisten Fällen wird gar nicht daran interessiert, alle möglichen Verbindungen zwischen allen Punkten zu berechnen, das wäre nämlich viel zu aufwendig, sondern wir wollen nur zwischen zwei gegebenen Punkten eine Verbindung finden. Abbildung 3.31 zeigt die Implementierung einer Prozedur `reachable(x, y, r)`, die überprüft, ob es in dem Graphen r eine Verbindung von x nach y gibt und die diese Verbindung berechnet. Das vollständige Programm finden Sie in der Datei `find-path.stlx`. Wir diskutieren nun die Implementierung der Prozedur `reachable`.

1. In Zeile 2 initialisieren wir p so, dass zunächst nur der Pfad der Länge 0, der mit dem Punkt x startet, in p liegt.
2. In Zeile 6 selektieren wir die Pfade aus p , die zum Ziel y führen.

3. Wenn wir dann in Zeile 7 feststellen, dass wir einen solchen Pfad berechnet haben, geben wir einen dieser Pfade in Zeile 8 zurück.
4. Falls es nicht gelingt einen solchen Pfad zu berechnen und wir keine neuen Pfade mehr finden können, verlassen wir die Prozedur in Zeile 11 mit dem Befehl **return**. Da wir bei diesem **return**-Befehl keinen Wert zurückgeben, ist der Rückgabewert der Prozedur in diesem Fall automatisch Ω .

```

1  reachable := procedure(x, y, r) {
2      p := { [x] };
3      while (true) {
4          oldP := p;
5          p := p + pathProduct(p, r);
6          found := { l in p | l[#1] == y };
7          if (found != {}) {
8              return arb(found);
9          }
10         if (p == oldP) {
11             return;
12         }
13     }
14 };

```

Abbildung 3.31: Berechnung aller Verbindungen zwischen zwei Punkten

3.8.3 Der Bauer mit dem Wolf, der Ziege und dem Kohl

Wir präsentieren nun eine betriebswissenschaftliche Anwendung des oben entwickelten Algorithmus und betrachten folgendes Problem.

Ein Agrarökonom will mit einem Wolf, einer Ziege und einem Kohl über einen Fluß übersetzen, um diese als Waren auf dem Markt zu verkaufen. Das Boot ist aber so klein, dass er nicht mehr als zwei Waren gleichzeitig mitnehmen kann. Wenn der Bauer den Wolf mit der Ziege allein läßt, dann frißt der Wolf die Ziege und wenn er die Ziege mit dem Kohl allein läßt, dann frißt die Ziege den Kohl.

Wir wollen einen Fahrplan entwickeln, mit dem der Agrarökonom alle seine Waren unbeschadet zum Markt bringen kann. Dazu modellieren wir das Rätsel als Erreichbarkeits-Problem in einem Graphen. Die Punkte des Graphen beschreiben dabei die einzelnen Situationen, die auftreten können. Wir definieren eine Menge

$$\mathbf{all} := \{\text{"Bauer"}, \text{"Wolf"}, \text{"Ziege"}, \text{"Kohl"}\}.$$

Die einzelnen Punkte sind dann Paare von Mengen, haben also die Form

$$\langle s_1, s_2 \rangle \quad \text{mit } s_1, s_2 \subseteq \mathbf{all}.$$

Dabei gibt die Menge s_1 an, was am linken Ufer ist und s_2 gibt an, was am rechten Ufer ist. Die Menge aller Punkte können wir dann definieren als

$$p := \{ \langle s_1, s_2 \rangle \in 2^{\mathbf{all}} \times 2^{\mathbf{all}} \mid s_1 \cup s_2 = \mathbf{all} \wedge s_1 \cap s_2 = \{\} \}.$$

Die Bedingung $s_1 \cup s_2 = \mathbf{all}$ stellt dabei sicher, dass nichts verloren geht: Jedes der Elemente aus \mathbf{all} ist entweder am linken oder am rechten Ufer. Die Bedingung $s_1 \cap s_2 = \{\}$ verbietet die Bilokalisation von Objekten, sie stellt also sicher, dass kein Element aus der Menge \mathbf{all} gleichzeitig am linken und am rechten Ufer ist.

Als nächstes definieren wir den Graphen r , also die möglichen Verbindungen zwischen Punkten. Dazu definieren wir eine Prozedur $problem(s)$. Hierbei ist s eine Menge von Objekten, die an einem Ufer sind. Die

Prozedur $problem(s)$ liefert genau dann **true**, wenn es bei der Menge s ein Problem gibt, weil entweder die Ziege mit dem Kohl allein ist, oder aber der Wolf die Ziege frißt.

```
problem := procedure(s) {
  return "goat" in s && "cabbage" in s || "wolf" in s && "goat" in s;
};
```

Damit können wir eine Relation r_1 wie folgt definieren:

$$r_1 := \left\{ \langle \langle s_1, s_2 \rangle, \langle s_1 \setminus b, s_2 \cup b \rangle \rangle \mid \langle s_1, s_2 \rangle \in P \wedge b \subseteq s_1 \wedge \text{"Bauer"} \in b \wedge card(b) \leq 2 \wedge \neg problem(s_1 \setminus b) \right\}.$$

Diese Menge beschreibt alle die Fahrten, bei denen der Bauer vom linken Ufer zum rechten Ufer fährt und bei denen zusätzlich sichergestellt ist, dass am linken Ufer nach der Überfahrt kein Problem auftritt. Die einzelnen Terme werden wie folgt interpretiert:

1. $\langle s_1, s_2 \rangle$ ist der Zustand vor der Überfahrt des Bootes, s_1 gibt also die Objekte am linken Ufer an, s_2 sind die Objekte am rechten Ufer.
2. b ist der Inhalt des Bootes, daher beschreibt $\langle s_1 \setminus b, s_2 \cup b \rangle$ den Zustand nach der Überfahrt des Bootes: Links sind nun nur noch die Objekte aus $s_1 \setminus b$, dafür sind rechts dann die Objekte $s_2 \cup b$.
Die Bedingungen lassen sich wie folgt interpretieren.
3. $b \subseteq s_1$: Es können natürlich nur solche Objekte ins Boot genommen werden, die vorher am linken Ufer waren.
4. $\text{"Bauer"} \in b$: Der Bauer muß auf jeden Fall ins Boot, denn weder der Wolf noch die Ziege können rudern.
5. $card(b) \leq 2$: Die Menge der Objekte im Boot darf nicht mehr als zwei Elemente haben, denn im Boot ist nur für zwei Platz.
6. $\neg problem(s_1 \setminus b)$: Am linken Ufer soll es nach der Überfahrt kein Problem geben, denn der Bauer ist ja hinterher am rechten Ufer.

In analoger Weise definieren wir nun eine Relation r_2 die die Überfahrten vom rechten Ufer zum linken Ufer beschreibt:

$$r_2 := \left\{ \langle \langle s_1, s_2 \rangle, \langle s_1 \cup b, s_2 \setminus b \rangle \rangle \mid \langle s_1, s_2 \rangle \in P \wedge b \subseteq s_2 \wedge \text{"Bauer"} \in b \wedge card(b) \leq 2 \wedge \neg problem(s_2 \setminus b) \right\}.$$

Die gesamte Relation r definieren wir nun als

$$r := r_1 \cup r_2.$$

Als nächstes müssen wir den Start-Zustand modellieren. Am Anfang sind alle am linken Ufer, also wird der Start-Zustand durch das Paar

$$\langle \{\text{"Bauer"}, \text{"Wolf"}, \text{"Ziege"}, \text{"Kohl"}\}, \{\} \rangle.$$

Beim Ziel ist es genau umgekehrt, dann sollen alle auf der rechten Seite des Ufers sein:

$$\langle \{\}, \{\text{"Bauer"}, \text{"Wolf"}, \text{"Ziege"}, \text{"Kohl"}\} \rangle.$$

Damit haben wir das Problem in der Mengenlehre modelliert und können die im letzten Abschnitt entwickelte Prozedur **reachable** benutzen, um das Problem zu lösen. In der Abbildung 3.32 finden Sie das Programm **wolf-goat-cabbage.stlx**, in dem die oben ausgeführten Überlegungen in SETLX umgesetzt wurden. Die von diesem Programm berechnete Lösung finden Sie in Abbildung 3.33.

```

1  all := { "farmer", "wolf", "goat", "cabbage" };
2  p   := pow(all);
3  r1  := { [ s, s - b ] : s in p, b in pow(s)
4          | "farmer" in b && #b <= 2 && !problem(s - b)
5          };
6  r2  := { [ s, s + b ] : s in p, b in pow(all - s)
7          | "farmer" in b && #b <= 2 && !problem(all - (s + b))
8          };
9  r    := r1 + r2;
10 start := [ all, {} ];
11 goal  := [ {}, all ];
12 path  := findPath(start, goal, r);

```

Abbildung 3.32: Wie kommt der Bauer ans andere Ufer?

```

1  {"Kohl", "Ziege", "Wolf", "Bauer"} {}
2                                     >> {"Ziege", "Bauer"} >>
3  {"Kohl", "Wolf"} {} {"Ziege", "Bauer"}
4                                     << {"Bauer"} <<<<
5  {"Kohl", "Wolf", "Bauer"} {} {"Ziege"}
6                                     >> {"Wolf", "Bauer"} >>>
7  {"Kohl"} {} {"Ziege", "Wolf", "Bauer"}
8                                     << {"Ziege", "Bauer"} <<
9  {"Kohl", "Ziege", "Bauer"} {} {"Wolf"}
10                                    >> {"Kohl", "Bauer"} >>>
11 {"Ziege"} {} {"Kohl", "Wolf", "Bauer"}
12                                    << {"Bauer"} <<<<
13 {"Ziege", "Bauer"} {} {"Kohl", "Wolf"}
14                                    >> {"Ziege", "Bauer"} >>
15 {} {} {"Kohl", "Ziege", "Wolf", "Bauer"}

```

Abbildung 3.33: Ein Fahrplan für den Bauern

3.9 Terme und Matching

Neben den bisher vorgestellten Datenstrukturen gibt es noch eine weitere wichtige Datenstruktur, die sogenannten *Terme*, die insbesondere nützlich ist, wenn wir *symbolische Programme* schreiben wollen. Darunter verstehen wir solche Programme, die Formeln manipulieren. Wollen wir beispielsweise ein Programm schreiben, dass als Eingabe einen String wie

`"x * sin(x)"`

einliest, diesen String als eine Funktion in der Variablen `"x"` interpretiert und dann die Ableitung dieser Funktion nach der Variablen `"x"` berechnet, so sprechen wir von einem *symbolischen Programm*. Wollen wir einen Ausdruck wie `"x * sin(x)"` darstellen, so eignen sich *Terme* am besten dazu. Im nächsten Unter-Abschnitt werden wir zunächst *Terme* zusammen mit den in SETLX vordefinierten Funktionen vorstellen, die zur Verarbeitung von Termen benutzt werden können. Anschließend stellen wir das sogenannte *Matching* vor, mit dessen Hilfe sich Terme besonders leicht manipulieren lassen.

3.9.1 Konstruktion und Manipulation von Termen

Terme werden mit Hilfe sogenannter *Funktions-Zeichen* gebildet. Es ist wichtig, dass Sie Funktions-Zeichen nicht mit Funktionen oder Variablen verwechseln. In SETLX beginnen Funktionen-Zeichen im Gegensatz zu einem Variablen-Namen daher mit einem großen Buchstaben. Auf den Großbuchstaben können dann beliebig viele Buchstaben, Ziffern und der Unterstrich `"_"` folgen. Zusätzlich gibt es noch Funktionszeichen, die mit dem Zeichen `"^"` beginnen. Solche Funktions-Zeichen werden intern von SETLX verwendet um Operator-Symbole wie `"+"` oder `"*"` darzustellen. Die folgenden Strings können beispielsweise als Funktions-Zeichen verwendet werden:

`F`, `FabcXYZ`, `^sum`, `Hugo_`.

Damit sind wir nun in der Lage, Term zu definieren. Ist F ein Funktions-Zeichen und sind t_1, t_2, \dots , beliebige SETLX-Werte, so ist der Ausdruck

$F(t_1, t_2, \dots, t_n)$

ein Term. Beachten Sie, dass Terme ganz ähnlich aussehen wie die Aufrufe von Funktionen. Terme und Aufrufe von Funktionen unterscheiden sich nur dadurch, dass bei einem Term links vor der ersten öffnenden Klammer ein Funktions-Zeichen steht, während bei einem Funktions-Aufruf dort statt dessen eine Variable steht, der eine Funktions-Definition zugewiesen worden ist.

Beispiele:

1. `Adresse("Rotebühlplatz 41", 70178, "Stuttgart")`

ist ein Term, der eine Adresse repräsentiert.

2. `Product(Variable("x"), Sin(Variable("x")))`

ist ein Term, der einen arithmetischen Ausdruck repräsentiert, den Sie mathematisch als $x \cdot \sin(x)$ schreiben würden. ◇

An dieser Stelle fragen Sie sich vielleicht, wie Terme ausgewertet werden. Die Antwort ist: **Gar nicht!** Terme werden nur dazu benutzt, Daten darzustellen. Terme sind also bereits Werte genauso wie auch Zahlen, Strings, Mengen oder Listen als Werte aufgefasst werden. Genausowenig wie Sie die Zahl 42 auswerten müssen, müssen Sie einen Term auswerten.

Nehmen wir einmal an, dass es in SETLX keine Listen geben würde. Dann könnten wir Listen als Terme darstellen. Zunächst würden wir ein Funktions-Zeichen benötigen, mit dem wir die leere Liste darstellen könnten. Wir wählen dazu das Funktions-Zeichen `Nil`. Damit haben wir dann also die Entsprechung

`Nil()` $\hat{=}$ `[]`.

Beachten Sie hier, dass die Klammern hinter dem Funktions-Zeichen `Nil` nicht weggelassen werden dürfen!

Um nun eine Liste darzustellen, deren erstes Element x ist und deren restliche Elemente durch die Restliste r gegeben sind, verwenden wir das Funktions-Zeichen **Cons**. Dann haben wir die Entsprechung

$$\text{Cons}(x, r) \hat{=} [x] + r.$$

Konkret können wir nun die Liste `[1,2,3]` durch den Term

$$\text{Cons}(1, \text{Cons}(2, \text{Cons}(3, \text{Nil}()))))$$

darstellen. In der Sprache *Prolog*, die wir später noch besprechen werden, werden Listen intern übrigens in ähnlicher Form als Terme dargestellt.

Es gibt zwei vordefinierte Funktionen in SETLX, mit denen wir auf die Komponenten eines Terms zugreifen können und es gibt eine Funktion, mit deren Hilfe wir Terme konstruieren können.

1. Die Funktion **fct** berechnet das Funktions-Zeichen eines Terms. Falls t ein Term der Form $F(s_1, \dots, s_n)$ ist, so ist das Ergebnis des Funktions-Aufrufs

$$\text{fct}(F(s_1, \dots, s_n))$$

das Funktions-Zeichen F dieses Terms. Beispielsweise liefert der Ausdruck

$$\text{fct}(\text{Cons}(1, \text{Cons}(2, \text{Cons}(3, \text{Nil}()))))$$

als Ergebnis das Funktions-Zeichen **"Cons"**.

2. Die Funktion **args** berechnet die Argumente eines Terms. Falls t ein Term der Form $F(s_1, \dots, s_n)$ ist, dann liefert der Ausdruck

$$\text{args}(F(s_1, \dots, s_n))$$

als Ergebnis die Liste $[s_1, \dots, s_n]$ der Argumente des Terms t . Beispielsweise liefert der Ausdruck

$$\text{args}(\text{Cons}(1, \text{Cons}(2, \text{Cons}(3, \text{Nil}()))))$$

das Ergebnis

$$[1, \text{Cons}(2, \text{Cons}(3, \text{Nil}()))].$$

3. Ist ein Funktions-Zeichen f und eine Liste l von Argumenten gegeben, so erzeugt die Funktion **makeTerm** durch den Aufruf

$$\text{makeTerm}(f, l)$$

einen Term t mit dem Funktions-Zeichen f und der Argument-Liste l , für t gilt also

$$\text{fct}(t) = f \quad \text{und} \quad \text{args}(t) = l.$$

Beispielsweise liefert der Aufruf

$$\text{makeTerm}(\text{"Cons"}, [1, \text{Nil}()])$$

als Ergebnis den Term

$$\text{Cons}(1, \text{Nil}()).$$

Diesen Term hätten wir natürlich auch unmittelbar hinschreiben können.

In Abbildung 3.34 auf Seite 68 sehen Sie die Implementierung einer Funktion **append**, deren Aufgabe es ist, ein Element x am Ende einer Liste l einzufügen, wobei vorausgesetzt ist, dass die Liste l als Term mit Hilfe der Funktions-Zeichen **"Cons"** und **"Nil"** dargestellt wird.

1. Zunächst wird in Zeile 2 überprüft, ob die Liste l leer ist. Die Liste l ist genau dann leer, wenn $l = \text{Nil}()$ gilt. Daher können wir einfach das Funktions-Zeichen von dem Term l testen um herauszufinden, ob l die leere Liste repräsentiert.
2. Falls l nicht leer ist, muss l die Form

```

1  append := procedure(l, x) {
2      if (fct(l) == "Nil") {
3          return Cons(x, Nil());
4      }
5      [head, tail] := args(l);
6      return Cons(head, append(tail, x));
7  };

```

Abbildung 3.34: Einfügen eines Elements am Ende einer Liste.

$l = \text{Cons}(\text{head}, \text{tail})$

haben. Dann ist *head* das erste Element der Liste *l* und *tail* bezeichnet die Liste der restlichen Elemente. In diesem Fall müssen wir *x* rekursiv in die Liste *tail* einfügen. Als Ergebnis wird in Zeile 6 dann eine neue Liste erzeugt, deren erstes Element *head* ist, während die Liste der restlichen Elemente durch den rekursiven Aufruf von *append* berechnet wird.

In manchen Fällen ist es sehr unbequem, dass Funktions-Zeichen immer mit einem großen Buchstaben beginnen müssen. Deswegen gibt es in SETLX einen Escape-Mechanismus, der es erlaubt, auch Funktionszeichen zu verwenden, die mit einem kleinen Buchstaben beginnen: Falls wir einem Funktionszeichen den Operator “@” voranstellen, dann darf das Funktionszeichen auch mit einem kleinen Buchstaben beginnen. Wollen wir beispielsweise Terme benutzen um algebraische Ausdrücke darzustellen, die trigonometrische Funktionen enthalten, so können wir einen Ausdruck der Form $\sin(x)$ durch den Term

`@sin("x")`

darstellen.

3.9.2 Matching

Der Umgang mit Termen wäre sehr mühsam, wenn wir die Terme jedesmal mit Hilfe der Funktionen *fct* und *args* auseinander nehmen müssten. Abbildung 3.35 zeigt eine weitere Implementierung der Funktion *append*, bei der wir die Kontroll-Struktur *match* an Stelle der Funktionen “*fct*” und “*args*” verwendet haben. In Zeile 3 wird überprüft, ob die Liste *l* leer ist. Die wahre Stärke des Matchings sehen wir allerdings in Zeile 4, denn dort wird nicht nur überprüft, ob die Liste *l* die Form

$\text{Cons}(\text{head}, \text{tail})$

hat, sondern gleichzeitig werden die Variablen *head* and *tail* so gesetzt, dass anschließend die Gleichung

$l = \text{Cons}(\text{head}, \text{tail})$

erfüllt ist.

```

1  append := procedure(l, x) {
2      match (l) {
3          case Nil():          return Cons(x, Nil());
4          case Cons(head, tail): return Cons(head, append(tail, x));
5      }
6  };

```

Abbildung 3.35: Implementierung von *append* mit Hilfe von *Matching*.

Im Allgemeinen ist ein *match*-Block so ähnlich aufgebaut wie ein *switch*-Block und hat die in Abbildung 3.36 gezeigte Struktur. Hier bezeichnet *e* einen Ausdruck, dessen Auswertung einen Term ergibt. Die Ausdrücke

t_1, \dots, t_n sind sogenannte *Muster*, die freie Variablen enthalten. Bei der Auswertung eines **Match-Blocks** versucht SETLX die in dem Muster t_i auftretenden Variablen so zu setzen, dass das Muster zu dem Ergebnis der Auswertung von e gleich ist. Gelingt dies, so wird die mit $body_i$ bezeichnete Gruppe von Befehlen ausgeführt. Andernfalls versucht SETLX das nächste Muster t_{i+1} mit e zur Deckung zu bringen. Falls keines der Muster t_1, \dots, t_n mit e zur Deckung zu bringen ist, wird ersatzweise $body_{n+1}$ ausgeführt.

```

1  match (e) {
2      case  $t_1$  :  $body_1$ 
3          :
4      case  $t_n$  :  $body_n$ 
5      default:  $body_{n+1}$ 
6  }
```

Abbildung 3.36: Struktur eines Match-Blocks

```

1  diff := procedure(t, x) {
2      match (t) {
3          case  $t_1 + t_2$  :
4              return  $\text{diff}(t_1, x) + \text{diff}(t_2, x)$ ;
5          case  $t_1 - t_2$  :
6              return  $\text{diff}(t_1, x) - \text{diff}(t_2, x)$ ;
7          case  $t_1 * t_2$  :
8              return  $\text{diff}(t_1, x) * t_2 + t_1 * \text{diff}(t_2, x)$ ;
9          case  $t_1 / t_2$  :
10             return (  $\text{diff}(t_1, x) * t_2 - t_1 * \text{diff}(t_2, x)$  ) /  $t_2 * t_2$ ;
11         case  $f ** g$  :
12             return  $\text{diff}(\text{@exp}(g * \text{@ln}(f)), x)$ ;
13         case  $\text{ln}(a)$  :
14             return  $\text{diff}(a, x) / a$ ;
15         case  $\text{exp}(a)$  :
16             return  $\text{diff}(a, x) * \text{@exp}(a)$ ;
17         case  $\sim\text{variable}(x)$  : // x is defined above as second argument
18             return 1;
19         case  $\sim\text{variable}(y)$  : // y not yet defined, matches any other variable
20             return 0;
21         case  $n \mid \text{isNumber}(n)$ :
22             return 0;
23     }
24 }
```

Abbildung 3.37: A function to perform symbolic differentiation.

Wir zeigen zum Abschluss dieses Abschnitts ein komplexeres Beispiel. Die in Abbildung 3.37 auf Seite 69 gezeigte Funktion **diff** wird mit zwei Argumenten aufgerufen:

1. Das erste Argument t ist ein Term, der einen arithmetischen Ausdruck repräsentiert.
2. Das zweite Argument x ist ein String, der als Variable interpretiert wird.

Die Aufgabe der Funktion **diff** besteht darin, den durch t gegebenen Ausdruck nach der in x angegebenen Variablen zu differenzieren. Wollen wir beispielsweise die Funktion

$$x \mapsto x^x$$

nach x ableiten, so können wir die Funktion `diff` wie folgt aufrufen.

```
diff(parse("x ** x"), "x");
```

Hier wandelt die Funktion `parse` den String `"x ** x"` in einen Term um. Die genaue Struktur dieses Terms diskutieren wir weiter unten. Wir betrachten zunächst den `match`-Befehl in Abbildung 3.37. In Zeile 3 hat der zu differenzierende Ausdruck die Form `t1 + t2`. Um einen solchen Ausdruck nach einer Variablen x zu differenzieren, müssen wir sowohl `t1` als auch `t2` nach x differenzieren. Die dabei erhaltenen Ergebnisse sind dann zu addieren. Etwas interessanter ist Zeile 8, welche die Produkt-Regel der Differenzial-Rechnung umsetzt. Die Produkt-Regel lautet:

$$\frac{d}{dx}(t_1 \cdot t_2) = \frac{dt_1}{dx} \cdot t_2 + t_1 \cdot \frac{dt_2}{dx}.$$

Bemerken Sie, dass in Zeile 7 das Muster

```
t1 * t2
```

zum einen dazu dient, zu erkennen, dass der zu differenzierende Ausdruck ein Produkt ist, zum anderen aber auch die beiden Faktoren des Produkts extrahiert und an die Variablen t_1 und t_2 bindet. In den Zeilen 12 und 16 haben wir den Funktions-Zeichen `"exp"` und `"ln"` den Operator `"@"` vorangestellt müssen, denn sonst würden die Strings `"exp"` und `"ln"` nicht als Funktions-Zeichen sondern als Variablen aufgefasst werden.

Die Regel zur Berechnung der Ableitung eines Ausdrucks der Form f^g beruht auf der Gleichung

$$f^g = \exp(\ln(f^g)) = \exp(g \cdot \ln(f)),$$

die in Zeile 12 umgesetzt wird.

Um einen Ausdruck der Form $\ln(f)$ abzuleiten, müssen wir die Kettenregel anwenden. Da $\frac{d}{dx} \ln(x) = \frac{1}{x}$ ist, haben wir insgesamt

$$\frac{d}{dx} \ln(f) = \frac{1}{f} \cdot \frac{df}{dx}.$$

Diese Gleichung wurde in Zeile 14 verwendet. In analoger Weise wird dann in Zeile 16 mit Hilfe der Kettenregel ein Ausdruck der Form `exp(f)` abgeleitet.

Um das Beispiel in Abbildung 3.37 besser zu verstehen müssen wir wissen, wie die Funktion `parse` einen String in einen Term umwandelt. Die Funktion `parse` muss sowohl Operator-Symbole als auch Variablen verarbeiten. Eine Variable der Form `"x"` wird in den Term

```
^variable("x")
```

umgewandelt. Dies erklärt die Zeilen 19 und 21 von Abbildung 3.37.

Wir können die interne Darstellung eines Terms mit Hilfe der Funktion `"canonical"` ausgeben. Beispielsweise liefert der Ausdruck

```
canonical(parse("x ** x"))
```

das Ergebnis

```
^power(^variable("x"), ^variable("x")).
```

Dies zeigt, dass der Exponentiations-Operator `"**"` in SETLX intern durch das Funktions-Zeichen `"^power"` dargestellt wird. Die interne Darstellung des Operators `"+"` ist `"^sum"`, `"−"` wird durch das Funktions-Zeichen `"^difference"` dargestellt, `"*"` wird durch das Funktions-Zeichen `"^product"` dargestellt und der Operator `"/"` wird durch das Funktions-Zeichen `"^quotient"` dargestellt.

Terme sind in dem folgenden Sinne *viral*: Falls ein Argument eines der Operatoren `"+"`, `"−"`, `"*"`, `"/"`, `"\"` und `"%"` ein Term ist, so erzeugt der Operator als Ergebnis automatisch einen Term. Beispielsweise liefert der Ausdruck

```
parse("x") + 2
```

den Term

```
^sum(^variable("x"), 2).
```

Zeile 21 zeigt, dass an ein Muster in einem `case` eine Bedingung angeschlossen werden kann: Das Muster

```
case n:
```

passt zunächst auf jeden Term. Allerdings wollen wir in Zeile 21 nur Zahlen matchen. Daher haben wir an dieses Muster mit Hilfe des Operators “|” noch die Bedingung `isNumber(n)` angehängt, mit der wir sicherstellen, dass n tatsächlich eine Zahl ist.

3.9.3 Ausblick

Wir konnten in diesem einführenden Kapitel nur einen Teil der Sprache SETLX behandeln. Einige weitere Features der Sprache SETLX werden wir noch in den folgenden Kapiteln diskutieren. Zusätzlich finden Sie weitere Informationen in dem Tutorial, das im Netz unter der Adresse

<http://wwwlehre.dhbw-stuttgart.de/stroetmann/SetlX/tutorial.pdf>

abgelegt ist.

Bemerkung: Die meisten der in diesem Abschnitt vorgestellten Algorithmen sind nicht effizient. Sie dienen nur dazu, die Begriffsbildungen aus der Mengenlehre konkret werden zu lassen. Die Entwicklung effizienter Algorithmen ist Gegenstand des zweiten Semesters.

Kapitel 4

Grenzen der Berechenbarkeit

In jeder Disziplin der Wissenschaft wird die Frage gestellt, welche Grenzen die verwendeten Methoden haben. Wir wollen daher in diesem Kapitel beispielhaft ein Problem untersuchen, bei dem die Informatik an ihre Grenzen stößt. Es handelt sich um das Halte-Problem.

4.1 Das Halte-Problem

Das Halte-Problem ist die Frage, ob eine gegebene Funktion für eine bestimmte Eingabe terminiert. Bevor wir formal beweisen, dass das Halte-Problem im Allgemeinen unlösbar ist, wollen wir versuchen, anschaulich zu verstehen, warum dieses Problem schwer sein muss. Dieser informalen Betrachtung des Halte-Problems ist der nächste Abschnitt gewidmet. Im Anschluss an diesen Abschluss zeigen wir dann die Unlösbarkeit des Halte-Problems.

4.1.1 Informale Betrachtungen zum Halte-Problem

```
1  legendre := procedure(n) {
2      k := n * n + 1;
3      while (k < (n + 1) ** 2) {
4          if (isPrime(k)) { return true; }
5          k += 1;
6      }
7      return false;
8  };
9  findCounterExample := procedure(n) {
10     while (true) {
11         if (legendre(n)) {
12             n := n + 1;
13         } else {
14             print("Legendre was wrong, no prime between $n**2$ and $(n+1)**2$!");
15             return;
16         }
17     }
18  };
```

Abbildung 4.1: Eine Funktion zur Überprüfung der Vermutung von Legendre.

Um zu verstehen, warum das Halte-Problem schwer ist, betrachten wir das in Abbildung 4.1 gezeigte Programm. Dieses Programm ist dazu gedacht, die *Legendre Vermutung* zu überprüfen. Der französische Mathematiker Adrien-Marie Legendre (1752 — 1833) hatte vor etwa 200 Jahren die Vermutung ausgesprochen, dass zwischen zwei positiven Quadratzahlen immer eine Primzahl liegt. Die Frage, ob diese Vermutung richtig ist, ist auch Heute noch unbeantwortet. Die in Abbildung 4.1 definierte Funktion `legendre(n)` überprüft für eine gegebene positive natürliche Zahl n , ob zwischen n^2 und $(n+1)^2$ eine Primzahl liegt. Falls dies, wie von Legendre vorhergesagt, der Fall ist, gibt die Funktion als Ergebnis `true` zurück, andernfalls wird `false` zurück gegeben.

Abbildung 4.1 enthält darüber hinaus die Definition der Funktion `findCounterExample(n)`, die versucht, für eine gegebene positive natürliche Zahl n eine Zahl $k \geq n$ zu finden, so dass zwischen k^2 und $(k+1)^2$ keine Primzahl liegt. Die Idee bei der Implementierung dieser Funktion ist einfach: Zunächst überprüfen wir durch den Aufruf `legendre(n)`, ob zwischen n^2 und $(n+1)^2$ eine Primzahl ist. Falls dies der Fall ist, untersuchen wir anschließend das Intervall von $(n+1)^2$ bis $(n+2)^2$, dann das Intervall von $(n+2)^2$ bis $(n+3)^2$ und so weiter, bis wir schließlich eine Zahl k finden, so dass zwischen k^2 und $(k+1)^2$ keine Primzahl liegt. Falls Legendre Recht hatte, werden wir nie ein solches k finden und in diesem Fall wird der Aufruf `findCounterExample(1)` nicht terminieren.

Nehmen wir nun an, wir hätten ein schlaues Programm, nennen wir es `stops`, das als Eingabe eine SETLX Funktion f und ein Argument a verarbeitet und dass uns die Frage, ob die Berechnung von $f(a)$ terminiert, beantworten kann. Die Idee wäre im wesentlichen, dass gilt:

$$\text{stops}(f, a) = 1 \quad \text{g.d.w.} \quad \text{der Aufruf } f(a) \text{ terminiert.}$$

Falls der Aufruf $f(a)$ nicht terminiert, sollte statt dessen `stops(f, a) = 0` gelten. Wenn wir eine solche Funktion `stops` hätten, dann könnten wir

$$\text{stops}(\text{findCounterExample}, 1)$$

aufrufen und wüssten anschließend, ob die Vermutung von Legendre wahr ist oder nicht: Wenn

$$\text{stops}(\text{findCounterExample}, 1) = 1$$

ist, dann würde das heißen, dass der Funktions-Aufruf `findCounterExample(1)` terminiert. Das passiert aber nur dann, wenn ein Gegenbeispiel gefunden wird. Würde der Aufruf `stops(findCounterExample, 1)` statt dessen eine 0 zurück liefern, so könnten wir schließen, dass der Aufruf `findCounterExample(1)` nicht terminiert. Mithin würde die Funktion `findCounterExample` kein Gegenbeispiel finden und das würde heißen, dass die Vermutung von Legendre stimmt.

Es gibt eine Reihe weiterer offener mathematischer Probleme, die alle auf die Frage abgebildet werden können, ob eine gegebene Funktion terminiert. Daher zeigen die vorhergehenden Überlegungen, dass es sehr nützlich wäre, eine Funktion wie `stops` zur Verfügung zu haben. Andererseits können wir an dieser Stelle schon ahnen, dass die Implementierung der Funktion `stops` zumindest sehr schwierig wird. Im Gegensatz zu Herrn Legendre, der seine Vermutung nicht bewiesen hat, werden wir unsere Vermutung im nächsten Abschnitt formal begründen.

4.1.2 Formale Analyse des Halte-Problems

Wir werden zeigen, dass das Halte-Problem nicht durch ein Programm gelöst werden kann. Dazu führen wir folgende Definition ein.

Definition 16 (Test-Funktion) Ein String t ist eine *Test-Funktion* wenn t die Form

$$\text{procedure}(x) \{ \dots \}$$

hat und sich als SETLX-Funktion parsen läßt. Die Menge der Test-Funktionen bezeichnen wir mit TF . □

Beispiele:

1. $s_1 = \text{"procedure}(x) \{ \text{return } 0; \}"$
 s_1 ist eine (sehr einfache) Test-Funktion.

2. $s_2 = \text{"procedure}(x) \{ \text{while } (\text{true}) \{ x := x + 1; \} \}"$

s_2 ist eine Test-Funktion.

3. $s_3 = \text{"procedure}(x) \{ \text{return } ++x; \}"$

s_3 ist keine Test-Funktion, denn da SETLX den Präfix-Operator “++” nicht unterstützt, läßt sich der String s_3 nicht fehlerfrei parsen.

Um das Halte-Problem übersichtlicher formulieren zu können, führen wir noch drei zusätzliche Notationen ein.

Notation 17 (\rightsquigarrow , \downarrow , \uparrow) Ist t eine SETLX-Funktion, die k Argumente verarbeitet und sind a_1, \dots, a_k Argumente, so schreiben wir

$$t(a_1, \dots, a_k) \rightsquigarrow r$$

wenn der Aufruf $t(a_1, \dots, a_k)$ das Ergebnis r liefert. Sind wir an dem Ergebnis selbst nicht interessiert, sondern wollen nur angeben, daß ein Ergebnis existiert, so schreiben wir

$$t(a_1, \dots, a_k) \downarrow$$

und sagen, dass der Aufruf $t(a_1, \dots, a_k)$ *terminiert*. Terminiert der Aufruf $t(a_1, \dots, a_k)$ nicht, so schreiben wir

$$t(a_1, \dots, a_k) \uparrow$$

und sagen, dass der Aufruf $t(a_1, \dots, a_k)$ *divergiert*. □

Beispiele: Legen wir die Funktions-Definitionen zugrunde, die wir im Anschluß an die Definition des Begriffs der Test-Funktion gegeben haben, so gilt:

1. $\text{procedure}(x) \{ \text{return } 0; \}(\text{"emil"}) \rightsquigarrow 0$

2. $\text{procedure}(x) \{ \text{return } 0; \}(\text{"emil"}) \downarrow$

3. $\text{procedure}(x) \{ \text{while } (\text{true}) \{ x := x + 1; \} \}(\text{"hugo"}) \uparrow$

Das *Halte-Problem* für SETLX-Funktionen ist die Frage, ob es eine SetlX-Funktion

$$\text{stops} := \text{procedure}(t, a) \{ \dots \}$$

gibt, die als Eingabe eine Testfunktion t und einen String a erhält und die folgende Eigenschaft hat:

1. $t \notin TF \Leftrightarrow \text{stops}(t, a) \rightsquigarrow 2$.

Der Aufruf $\text{stops}(t, a)$ liefert genau dann den Wert 2 zurück, wenn t keine Test-Funktion ist.

2. $t \in TF \wedge t(a) \downarrow \Leftrightarrow \text{stops}(t, a) \rightsquigarrow 1$.

Der Aufruf $\text{stops}(t, a)$ liefert genau dann den Wert 1 zurück, wenn t eine Test-Funktion ist und der Aufruf $t(a)$ terminiert.

3. $t \in TF \wedge t(a) \uparrow \Leftrightarrow \text{stops}(t, a) \rightsquigarrow 0$.

Der Aufruf $\text{stops}(t, a)$ liefert genau dann den Wert 0 zurück, wenn t eine Test-Funktion ist und der Aufruf $t(a)$ nicht terminiert.

Falls eine SETLX-Funktion stops mit den obigen Eigenschaften existiert, dann sagen wir, dass das Halte-Problem für SETLX entscheidbar ist.

Theorem 18 (Turing, 1936) Das Halte-Problem ist unentscheidbar.

Beweis: Zunächst eine Vorbemerkung. Um die Unentscheidbarkeit des Halte-Problems nachzuweisen, müssen wir zeigen, dass etwas, nämlich eine Funktion mit gewissen Eigenschaften nicht existiert. Wie kann so ein Beweis überhaupt funktionieren? Wie können wir überhaupt zeigen, dass irgendetwas nicht existiert? Die einzige Möglichkeit zu zeigen, dass etwas nicht existiert ist indirekt: Wir nehmen also an, dass eine Funktion `stops` existiert, die das Halte-Problem löst. Aus dieser Annahme werden wir einen Widerspruch ableiten. Dieser Widerspruch zeigt uns dann, dass eine Funktion `stops` mit den gewünschten Eigenschaften nicht existieren kann. Um zu einem Widerspruch zu kommen, definieren wir den String `turing` wie in Abbildung 4.2 gezeigt.

```

1  turing := "procedure(x) {
2      result := stops(x, x);
3      if (result == 1) {
4          while (true) {
5              print("... looping ...");
6          }
7      }
8      return result;
9  };"

```

Abbildung 4.2: Die Definition des Strings `turing`.

Mit dieser Definition ist klar, dass `turing` eine Test-Funktion ist:

$$\text{turing} \in TF.$$

Damit sind wir in der Lage, den String `turing` als Eingabe der Funktion `stops` zu verwenden. Wir betrachten nun den folgenden Aufruf:

`stops(turing, turing);`

Offenbar ist `turing` eine Test-Funktion. Daher können nur zwei Fälle auftreten:

$$\text{stops}(\text{turing}, \text{turing}) \rightsquigarrow 0 \quad \vee \quad \text{stops}(\text{turing}, \text{turing}) \rightsquigarrow 1.$$

Diese beiden Fälle analysieren wir nun im Detail:

1. $\text{stops}(\text{turing}, \text{turing}) \rightsquigarrow 0$.

Nach der Spezifikation von `stops` bedeutet dies

$$\text{turing}(\text{turing}) \uparrow$$

Schauen wir nun, was wirklich beim Aufruf `turing(turing)` passiert: In Zeile 2 erhält die Variable `result` den Wert 0 zugewiesen. In Zeile 3 wird dann getestet, ob `result` den Wert 1 hat. Dieser Test schlägt fehl. Daher wird der Block der `if`-Anweisung nicht ausgeführt und die Funktion liefert als nächstes in Zeile 8 den Wert 0 zurück. Insbesondere terminiert der Aufruf also, im Widerspruch zu dem, was die Funktion `stops` behauptet hat.

Damit ist der erste Fall ausgeschlossen.

2. $\text{stops}(\text{turing}, \text{turing}) \rightsquigarrow 1$.

Aus der Spezifikation der Funktion `stops` folgt, dass der Aufruf `turing(turing)` terminiert:

$$\text{turing}(\text{turing}) \downarrow$$

Schauen wir nun, was wirklich beim Aufruf `turing(turing)` passiert: In Zeile 2 erhält die Variable `result` den Wert 1 zugewiesen. In Zeile 3 wird dann getestet, ob `result` den Wert 1 hat. Diesmal gelingt der Test. Daher wird der Block der `if`-Anweisung ausgeführt. Dieser Block besteht aber nur aus einer Endlos-Schleife, aus der wir nie wieder zurück kommen. Das steht im Widerspruch zu dem, was die Funktion `stops` behauptet hat.

Damit ist der zweite Fall ausgeschlossen.

Insgesamt haben wir also in jedem Fall einen Widerspruch erhalten. Also ist die Annahme, dass die SETLX-Funktion `stops` das Halte-Problem löst, falsch. Insgesamt haben wir gezeigt, dass es keine SETLX-Funktion geben kann, die das Halte-Problem löst. \square

Bemerkung: Der Nachweis, dass das Halte-Problem unlösbar ist, wurde 1936 von Alan Turing (1912 – 1954) [Tur36] erbracht. Turing hat das Problem damals natürlich nicht für die Sprache SETLX gelöst, sondern für die heute nach ihm benannten *Turing-Maschinen*. Eine Turing-Maschine ist abstrakt gesehen nichts anderes als eine Beschreibung eines Algorithmus. Turing hat also gezeigt, dass es keinen Algorithmus gibt, der entscheiden kann, ob ein gegebener anderer Algorithmus terminiert.

Bemerkung: An dieser Stelle können wir uns fragen, ob es vielleicht eine andere Programmier-Sprache gibt, in der wir das Halte-Problem dann vielleicht doch lösen könnten. Wenn es in dieser Programmier-Sprache Unterprogramme gibt, und wenn wir dort Programm-Texte als Argumente von Funktionen übergeben können, dann ist leicht zu sehen, dass der obige Beweis der Unlösbarkeit des Halte-Problems sich durch geeignete syntaktische Modifikationen auch auf die andere Programmier-Sprache übertragen läßt.

4.2 Unlösbarkeit des Äquivalenz-Problems

Es gibt noch eine ganze Reihe anderer Funktionen, die nicht berechenbar sind. In der Regel werden wir den Nachweis, dass eine bestimmte Funktion nicht berechenbar ist, dadurch führen, dass wir zunächst annehmen, dass die gesuchte Funktion doch implementierbar ist. Unter dieser Annahme konstruieren wir dann eine Funktion, die das Halte-Problem löst, was im Widerspruch zu dem am Anfang dieses Abschnitts bewiesenen Sachverhalt steht. Dieser Widerspruch zwingt uns zu der Folgerung, dass die gesuchte Funktion nicht implementierbar ist. Wir werden dieses Verfahren an einem Beispiel demonstrieren. Vorweg benötigen wir aber noch eine Definition.

Definition 19 (\simeq) Es seien t_1 und t_2 zwei SETLX-Funktionen und a_1, \dots, a_k seien Argumente, mit denen wir diese Funktionen füttern können. Wir definieren

$$t_1(a_1, \dots, a_k) \simeq t_2(a_1, \dots, a_k)$$

g.d.w. einer der beiden folgenden Fälle auftritt:

1. $t_1(a_1, \dots, a_k) \uparrow \wedge t_2(a_1, \dots, a_k) \uparrow$
2. $\exists r : (t_1(a_1, \dots, a_k) \rightsquigarrow r \wedge t_2(a_1, \dots, a_k) \rightsquigarrow r)$

In diesem Fall sagen wir, dass die beiden Funktions-Aufrufe $t_1(a_1, \dots, a_k) \simeq t_2(a_1, \dots, a_k)$ *partiell äquivalent* sind. \square

Wir kommen jetzt zum *Äquivalenz-Problem*. Die Funktion `equal`, die die Form

`equal := procedure(p1, p2, a) { ... }`

hat, möge folgender Spezifikation genügen:

1. $p_1 \notin TF \vee p_2 \notin TF \Leftrightarrow \text{equal}(p_1, p_2, a) \rightsquigarrow 2.$
2. Falls
 - (a) $p_1 \in TF$,
 - (b) $p_2 \in TF$ und
 - (c) $p_1(a) \simeq p_2(a)$

gilt, dann muß gelten:

$$\text{equal}(p_1, p_2, a) \rightsquigarrow 1.$$

3. Ansonsten gilt

$$\text{equal}(p_1, p_2, a) \rightsquigarrow 0.$$

Wir sagen, dass eine Funktion, die der eben angegebenen Spezifikation genügt, das *Äquivalenz-Problem* löst.

Theorem 20 (Rice, 1953) *Das Äquivalenz-Problem ist unlösbar.*

Beweis: Wir führen den Beweis indirekt und nehmen an, dass es doch eine Implementierung der Funktion `equal` gibt, die das Äquivalenz-Problem löst. Wir betrachten die in Abbildung 4.3 angegebene Implementierung der Funktion `stops`.

```
1  stops := procedure(p, a) {
2      f := "procedure(x) { while (true) { x := x + x; } }";
3      e := equal(f, p, a);
4      if (e == 2) {
5          return 2;
6      } else {
7          return 1 - e;
8      }
9  }
```

Abbildung 4.3: Eine Implementierung der Funktion `stops`.

Zu beachten ist, dass in Zeile 2 die Funktion `equal` mit einem String aufgerufen wird, der eine Test-Funktion ist. Diese Test-Funktion hat die folgende Form:

```
procedure(x) { while (true) { x := x + x; } };
```

Es ist offensichtlich, dass diese Funktion für kein Ergebnis terminiert. Ist also das Argument p eine Test-Funktion, so liefert die Funktion `equal` immer dann den Wert 1, wenn $p(a)$ nicht terminiert, andernfalls muss sie den Wert 0 zurück geben. Damit liefert die Funktion `stops` aber für eine Test-Funktion p und ein Argument a genau dann 1, wenn der Aufruf $p(a)$ terminiert und würde folglich das Halte-Problem lösen. Das kann nicht sein, also kann es keine Funktion `equal` geben, die das Äquivalenz-Problem löst. \square

Die Unlösbarkeit des Äquivalenz-Problems und vieler weiterer praktisch interessanter Problem folgen aus einem 1953 von Henry G. Rice [Ric53] bewiesenen Satz.

Kapitel 5

Aussagenlogik

5.1 Motivation

Die Aussagenlogik beschäftigt sich mit der Verknüpfung einfacher Aussagen durch *Junktoren*. Dabei sind Junktoren Worte wie “und”, “oder”, “nicht”, “wenn \dots , dann”, und “genau dann, wenn”. Einfache Aussagen sind dabei Sätze, die

- einen Tatbestand ausdrücken, der entweder wahr oder falsch ist und
- selber keine Junktoren enthalten.

Beispiele für einfache Aussagen sind

1. “Die Sonne scheint.”
2. “Es regnet.”
3. “Am Himmel ist ein Regenbogen.”

Einfache Aussagen dieser Art bezeichnen wir auch als *atomare* Aussagen, weil sie sich nicht weiter in Teilaussagen zerlegen lassen. Atomare Aussagen lassen sich mit Hilfe der eben angegebenen Junktoren zu *zusammengesetzten Aussagen* verknüpfen. Ein Beispiel für eine zusammengesetzte Aussage wäre

Wenn die Sonne scheint und es regnet, dann ist ein Regenbogen am Himmel. (1)

Die Aussage ist aus den drei atomaren Aussagen “Die Sonne scheint.”, “Es regnet.”, und “Am Himmel ist ein Regenbogen.” mit Hilfe der Junktoren “und” und “wenn \dots , dann” aufgebaut worden. Die Aussagenlogik untersucht, wie sich der Wahrheitswert zusammengesetzter Aussagen aus dem Wahrheitswert der einzelnen Teilaussagen berechnen läßt. Darauf aufbauend wird dann gefragt, in welcher Art und Weise wir aus gegebenen Aussagen neue Aussagen logisch folgern können.

Um die Struktur komplexerer Aussagen übersichtlich werden zu lassen, führen wir in der Aussagenlogik zunächst sogenannte *Aussage-Variablen* ein. Diese stehen für atomare Aussagen. Zusätzlich führen wir für die Junktoren “nicht”, “und”, “oder”, “wenn, \dots dann”, und “genau dann, wenn” die folgenden Abkürzungen ein:

1. $\neg a$ für nicht a
2. $a \wedge b$ für a und b
3. $a \vee b$ für a oder b
4. $a \rightarrow b$ für wenn a , dann b
5. $a \leftrightarrow b$ für a genau dann, wenn b

Aussagenlogische Formeln werden aus Aussage-Variablen mit Hilfe von Junktoren aufgebaut. Bestimmte aussagenlogische Formeln sind offenbar immer wahr, egal was wir für die einzelnen Teilaussagen einsetzen. Beispielsweise ist eine Formel der Art

$$p \vee \neg p$$

unabhängig von dem Wahrheitswert der Aussage p immer wahr. Eine aussagenlogische Formel, die immer wahr ist, bezeichnen wir als eine *Tautologie*. Andere aussagenlogische Formeln sind nie wahr, beispielsweise ist die Formel

$$p \wedge \neg p$$

immer falsch. Eine Formel heißt *erfüllbar*, wenn es wenigstens eine Möglichkeit gibt, bei der die Formel wahr wird. Im Rahmen der Vorlesung werden wir verschiedene Verfahren entwickeln, mit denen es möglich ist zu entscheiden, ob eine aussagenlogische Formel eine Tautologie ist oder ob Sie wenigstens erfüllbar ist. Solche Verfahren spielen in der Praxis eine wichtige Rolle.

5.2 Anwendungen der Aussagenlogik

Die Aussagenlogik bildet nicht nur die Grundlage für die Prädikatenlogik, sondern sie hat auch wichtige praktische Anwendungen. Aus der großen Zahl der industriellen Anwendungen möchte ich stellvertretend vier Beispiele nennen:

1. Analyse und Design digitaler Schaltungen.

Komplexe digitale Schaltungen bestehen heute aus mehreren Millionen logischen Gattern¹. Ein Gatter ist dabei, aus logischer Sicht betrachtet, ein Baustein, der einen der logischen Junktoren wie “und”, “oder”, “nicht”, etc. auf elektronischer Ebene repräsentiert.

Die Komplexität solcher Schaltungen wäre ohne den Einsatz rechnergestützter Verfahren zur Verifikation nicht mehr beherrschbar. Die dabei eingesetzten Verfahren sind Anwendungen der Aussagenlogik.

Eine ganz konkrete Anwendung ist der Schaltungs-Vergleich. Hier werden zwei digitale Schaltungen als aussagenlogische Formeln dargestellt. Anschließend wird versucht, mit aussagenlogischen Mitteln die Äquivalenz dieser Formeln zu zeigen. Software-Werkzeuge, die für die Verifikation digitaler Schaltungen eingesetzt werden, kosten heutzutage über 100 000 \$². Dies zeigt die wirtschaftliche Bedeutung der Aussagenlogik.

2. Erstellung von Einsatzplänen (*crew scheduling*).

International tätige Fluggesellschaften müssen bei der Einteilung Ihrer Crews einerseits gesetzlich vorgesehene Ruhezeiten einhalten, wollen aber ihr Personal möglichst effizient einsetzen. Das führt zu Problemen, die sich mit Hilfe aussagenlogischer Formeln beschreiben und lösen lassen.

3. Erstellung von Verschußplänen für die Weichen und Signale von Bahnhöfen.

Bei einem größeren Bahnhof gibt es einige hundert Weichen und Signale, die ständig neu eingestellt werden müssen, um sogenannte *Fahrstraßen* für die Züge zu realisieren. Verschiedene Fahrstraßen sollen sich natürlich nicht kreuzen. Die einzelnen Fahrstraßen werden durch sogenannte *Verschußpläne* beschrieben. Die Korrektheit solcher Verschußpläne kann durch aussagenlogische Formeln ausgedrückt werden.

4. Eine Reihe kombinatorischer Puzzles lassen sich als aussagenlogische Formeln kodieren und können dann mit Hilfe aussagenlogischer Methoden lösen. Als ein Beispiel werden wir in der Vorlesung das 8-Damen-Problem behandeln. Dabei geht es um die Frage, ob 8 Damen so auf einem Schachbrett angeordnet werden können, dass keine der Damen eine andere Dame bedroht.

¹ Die Version des Pentium 4 Prozessors mit dem *Northwood* Kernel enthält etwa 55 Millionen Transistoren.

² Die Firma Magma bietet beispielsweise den *Equivalence-Checker Quartz Formal* zum Preis von 150 000 \$ pro Lizenz an. Eine solche Lizenz ist dann drei Jahre lang gültig.

5.3 Formale Definition der aussagenlogischen Formeln

Wir behandeln zunächst die *Syntax* der Aussagenlogik und besprechen anschließend die *Semantik*. Die *Syntax* gibt an, wie Formeln geschrieben werden. Die *Semantik* befasst sich mit der Bedeutung der Formeln. Nachdem wir die Semantik der aussagenlogischen Formeln definiert haben, zeigen wir, wie sich diese Semantik in SETLX implementieren lässt.

5.3.1 Syntax der aussagenlogischen Formeln

Wir betrachten eine Menge \mathcal{P} von *Aussage-Variablen* als gegeben. Aussagenlogische Formeln sind dann Wörter, die aus dem Alphabet

$$\mathcal{A} := \mathcal{P} \cup \{\top, \perp, \neg, \vee, \wedge, \rightarrow, \leftrightarrow, (,)\}$$

gebildet werden. Wir definieren die Menge der aussagenlogischen Formeln \mathcal{F} durch eine induktive Definition:

1. $\top \in \mathcal{F}$ und $\perp \in \mathcal{F}$.

Hier steht \top für die Formel, die immer wahr ist, während \perp für die Formel steht, die immer falsch ist. Die Formel \top trägt auch den Namen *Verum*, für \perp sagen wir auch *Falsum*.

2. Ist $p \in \mathcal{P}$, so gilt auch $p \in \mathcal{F}$.

Jede aussagenlogische Variable ist also eine aussagenlogische Formel.

3. Ist $f \in \mathcal{F}$, so gilt auch $\neg f \in \mathcal{F}$.

4. Sind $f_1, f_2 \in \mathcal{F}$, so gilt auch

$$\begin{array}{ll} (f_1 \vee f_2) \in \mathcal{F} & (\text{gelesen: } f_1 \text{ oder } f_2), \\ (f_1 \wedge f_2) \in \mathcal{F} & (\text{gelesen: } f_1 \text{ und } f_2), \\ (f_1 \rightarrow f_2) \in \mathcal{F} & (\text{gelesen: } \text{wenn } f_1, \text{ dann } f_2), \\ (f_1 \leftrightarrow f_2) \in \mathcal{F} & (\text{gelesen: } f_1 \text{ genau dann, wenn } f_2). \end{array}$$

Die Menge \mathcal{F} ist nun die kleinste Teilmenge der aus dem Alphabet \mathcal{A} gebildeten Wörter, die den oben aufgestellten Forderungen genügt.

Beispiel: Gilt $\mathcal{P} = \{p, q, r\}$, so haben wir beispielsweise:

1. $p \in \mathcal{F}$,
2. $(p \wedge q) \in \mathcal{F}$,
3. $((\neg p \rightarrow q) \vee (q \rightarrow \neg p)) \in \mathcal{F}$. □

Um Klammern zu sparen, vereinbaren wir:

1. Äußere Klammern werden weggelassen, wir schreiben also beispielsweise

$$p \wedge q \quad \text{statt} \quad (p \wedge q).$$

2. Die Junktoren \vee und \wedge werden implizit links geklammert, d.h. wir schreiben

$$p \wedge q \wedge r \quad \text{statt} \quad (p \wedge q) \wedge r.$$

Operatoren, die implizit nach links geklammert werden, nennen wir *links-assoziativ*.

3. Der Junktor \rightarrow wird implizit rechts geklammert, d.h. wir schreiben

$$p \rightarrow q \rightarrow r \quad \text{statt} \quad p \rightarrow (q \rightarrow r).$$

Operatoren, die implizit nach rechts geklammert werden, nennen wir *rechts-assoziativ*.

4. Die Junktoren \vee und \wedge binden stärker als \rightarrow , wir schreiben also

$$p \wedge q \rightarrow r \quad \text{statt} \quad (p \wedge q) \rightarrow r$$

Beachten Sie, dass die Junktoren \wedge und \vee gleich stark binden. Dies ist anders als in der Sprache SETLX, denn dort bindet der Operator “&&” stärker als der Operator “||”.

5. Der Junktor \rightarrow bindet stärker als \leftrightarrow , wir schreiben also

$$p \rightarrow q \leftrightarrow r \quad \text{statt} \quad (p \rightarrow q) \leftrightarrow r.$$

5.3.2 Semantik der aussagenlogischen Formeln

Um aussagenlogischen Formeln einen Wahrheitswert zuordnen zu können, definieren wir zunächst die Menge \mathbb{B} der Wahrheitswerte:

$$\mathbb{B} := \{\text{true}, \text{false}\}.$$

Damit können wir nun den Begriff einer *aussagenlogischen Interpretation* festlegen.

Definition 21 (Aussagenlogische Interpretation) Eine *aussagenlogische Interpretation* ist eine Funktion

$$\mathcal{I} : \mathcal{P} \rightarrow \mathbb{B},$$

die jeder Aussage-Variablen $p \in \mathcal{P}$ einen Wahrheitswert $\mathcal{I}(p) \in \mathbb{B}$ zuordnet. □

Eine aussagenlogische Interpretation wird oft auch als *Belegung* der Aussage-Variablen mit Wahrheits-Werten bezeichnet.

Eine aussagenlogische Interpretation \mathcal{I} interpretiert die Aussage-Variablen. Um nicht nur Variablen sondern auch aussagenlogische Formel interpretieren zu können, benötigen wir eine Interpretation der Junktoren “ \neg ”, “ \wedge ”, “ \vee ”, “ \rightarrow ” und “ \leftrightarrow ”. Zu diesem Zweck definieren wir auf der Menge \mathbb{B} Funktionen \neg , \wedge , \vee , \rightarrow und \leftrightarrow mit deren Hilfe wir die aussagenlogischen Junktoren interpretieren können:

1. $\neg : \mathbb{B} \rightarrow \mathbb{B}$
2. $\wedge : \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$
3. $\vee : \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$
4. $\rightarrow : \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$
5. $\leftrightarrow : \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$

Wir haben in der Mengenlehre gesehen, dass Funktionen als spezielle Relationen aufgefaßt werden können. Die Funktion \neg dreht die Wahrheits-Werte um und kann daher als Relation wie folgt geschrieben werden:

$$\neg = \{\langle \text{true}, \text{false} \rangle, \langle \text{false}, \text{true} \rangle\}.$$

Es ist anschaulicher, die Funktionen durch eine Tabelle zu definieren. Diese Tabelle ist auf Seite 81 abgebildet.

p	q	$\neg(p)$	$\vee(p, q)$	$\wedge(p, q)$	$\rightarrow(p, q)$	$\leftrightarrow(p, q)$
true	true	false	true	true	true	true
true	false	false	true	false	false	false
false	true	true	true	false	true	false
false	false	true	false	false	true	true

Tabelle 5.1: Interpretation der Junktoren.

Nun können wir den Wert, den eine aussagenlogische Formel f unter einer aussagenlogischen Interpretation \mathcal{I} annimmt, durch Induktion nach dem Aufbau der Formel f definieren. Wir werden diesen Wert mit $\hat{\mathcal{I}}(f)$ bezeichnen. Wir setzen:

1. $\widehat{\mathcal{I}}(\perp) := \mathbf{false}$.
2. $\widehat{\mathcal{I}}(\top) := \mathbf{true}$.
3. $\widehat{\mathcal{I}}(p) := \mathcal{I}(p)$ für alle $p \in \mathcal{P}$.
4. $\widehat{\mathcal{I}}(\neg f) := \ominus(\widehat{\mathcal{I}}(f))$ für alle $f \in \mathcal{F}$.
5. $\widehat{\mathcal{I}}(f \wedge g) := \otimes(\widehat{\mathcal{I}}(f), \widehat{\mathcal{I}}(g))$ für alle $f, g \in \mathcal{F}$.
6. $\widehat{\mathcal{I}}(f \vee g) := \oslash(\widehat{\mathcal{I}}(f), \widehat{\mathcal{I}}(g))$ für alle $f, g \in \mathcal{F}$.
7. $\widehat{\mathcal{I}}(f \rightarrow g) := \ominus(\widehat{\mathcal{I}}(f), \widehat{\mathcal{I}}(g))$ für alle $f, g \in \mathcal{F}$.
8. $\widehat{\mathcal{I}}(f \leftrightarrow g) := \oplus(\widehat{\mathcal{I}}(f), \widehat{\mathcal{I}}(g))$ für alle $f, g \in \mathcal{F}$.

Um die Schreibweise nicht übermäßig kompliziert werden zu lassen, unterscheiden wir in Zukunft nicht mehr zwischen $\widehat{\mathcal{I}}$ und \mathcal{I} , wir werden das Hütchen über dem \mathcal{I} also weglassen.

Beispiel: Wir zeigen, wie sich der Wahrheits-Wert der Formel

$$(p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q$$

für die aussagenlogische Interpretation \mathcal{I} , die durch $\mathcal{I}(p) = \mathbf{true}$ und $\mathcal{I}(q) = \mathbf{false}$ definiert ist, berechnen läßt:

$$\begin{aligned}
\mathcal{I}\big((p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q\big) &= \ominus\big(\mathcal{I}((p \rightarrow q)), \mathcal{I}((\neg p \rightarrow q) \rightarrow q)\big) \\
&= \ominus\big(\ominus(\mathcal{I}(p), \mathcal{I}(q)), \mathcal{I}((\neg p \rightarrow q) \rightarrow q)\big) \\
&= \ominus\big(\ominus(\mathbf{true}, \mathbf{false}), \mathcal{I}((\neg p \rightarrow q) \rightarrow q)\big) \\
&= \ominus\big(\mathbf{false}, \mathcal{I}((\neg p \rightarrow q) \rightarrow q)\big) \\
&= \mathbf{true}
\end{aligned}$$

Beachten Sie, dass wir bei der Berechnung gerade so viele Teile der Formel ausgewertet haben, wie notwendig waren um den Wert der Formel zu bestimmen. Trotzdem ist die eben durchgeführte Rechnung für die Praxis zu umständlich. Stattdessen wird der Wert einer Formel direkt mit Hilfe der Tabelle 5.1 auf Seite 81 berechnet. Wir zeigen exemplarisch, wie wir den Wahrheits-Wert der Formel

$$(p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q$$

für beliebige Belegungen \mathcal{I} über diese Tabelle berechnen können. Um nun die Wahrheitswerte dieser Formel unter einer gegebenen Belegung der Aussage-Variablen bestimmen zu können, bauen wir eine Tabelle auf, die für jede in der Formel auftretende Teilformel eine Spalte enthält. Tabelle 5.2 auf Seite 82 zeigt die entstehende Tabelle.

p	q	$\neg p$	$p \rightarrow q$	$\neg p \rightarrow q$	$(\neg p \rightarrow q) \rightarrow q$	$(p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q$
true	true	false	true	true	true	true
true	false	false	false	true	false	true
false	true	true	true	true	true	true
false	false	true	true	false	true	true

Tabelle 5.2: Berechnung Der Wahrheitswerte von $(p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q$.

Betrachten wir die letzte Spalte der Tabelle so sehen wir, dass dort immer der Wert **true** auftritt. Also liefert die Auswertung der Formel $(p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q$ für jede aussagenlogische Belegung \mathcal{I} den Wert

true. Formeln, die immer wahr sind, haben in der Aussagenlogik eine besondere Bedeutung und werden als *Tautologien* bezeichnet.

Wir erläutern die Aufstellung dieser Tabelle anhand der zweiten Zeile. In dieser Zeile sind zunächst die aussagenlogischen Variablen p auf **true** und q auf **false** gesetzt. Bezeichnen wir die aussagenlogische Interpretation mit \mathcal{I} , so gilt also

$$\mathcal{I}(p) = \text{true} \text{ und } \mathcal{I}(q) = \text{false}.$$

Damit erhalten wir folgende Rechnung:

1. $\mathcal{I}(\neg p) = \ominus(\mathcal{I}(p)) = \ominus(\text{true}) = \text{false}$
2. $\mathcal{I}(p \rightarrow q) = \ominus(\mathcal{I}(p), \mathcal{I}(q)) = \ominus(\text{true}, \text{false}) = \text{false}$
3. $\mathcal{I}(\neg p \rightarrow q) = \ominus(\mathcal{I}(\neg p), \mathcal{I}(q)) = \ominus(\text{false}, \text{false}) = \text{true}$
4. $\mathcal{I}((\neg p \rightarrow q) \rightarrow q) = \ominus(\mathcal{I}(\neg p \rightarrow q), \mathcal{I}(q)) = \ominus(\text{true}, \text{false}) = \text{false}$
5. $\mathcal{I}((p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q) = \ominus(\mathcal{I}(p \rightarrow q), \mathcal{I}((\neg p \rightarrow q) \rightarrow q)) = \ominus(\text{false}, \text{false}) = \text{true}$

Für komplexe Formeln ist die Auswertung von Hand viel zu mühsam und fehleranfällig um praktikabel zu sein. Wir zeigen deshalb im übernächsten Abschnitt, wie sich dieser Prozeß automatisieren läßt.

5.3.3 Extensionale und intensionale Interpretationen der Aussagenlogik

Die Interpretation des aussagenlogischen Junktoren ist rein *extensional*: Wenn wir den Wahrheitswert der Formel

$$\mathcal{I}(f \rightarrow g)$$

berechnen wollen, so müssen wir die Details der Teilformeln f und g nicht kennen, es reicht, wenn wir die Werte $\mathcal{I}(f)$ und $\mathcal{I}(g)$ kennen. Das ist problematisch, denn in der Umgangssprache hat der Junktor “wenn \dots , dann” auch eine *kausale* Bedeutung. Mit der extensionalen Implikation wird der Satz

“Wenn $3 \cdot 3 = 8$, dann schneit es.”

als wahr interpretiert, denn die Formel $3 \cdot 3 = 8$ ist ja falsch. Das ist problematisch, weil wir diesen Satz in der Umgangssprache als sinnlos erkennen. Insofern ist die extensionale Interpretation des sprachlichen Junktors “wenn \dots , dann” nur eine Approximation der umgangssprachlichen Interpretation, die sich für die Mathematik und die Informatik aber als ausreichend erwiesen hat.

5.3.4 Implementierung in SetlX

Um die bisher eingeführten Begriffe nicht zu abstrakt werden zu lassen, entwickeln wir in SETLX ein Programm, mit dessen Hilfe sich Formeln auswerten lassen. Jedesmal, wenn wir ein Programm zur Berechnung irgendwelcher Wert entwickeln wollen, müssen wir uns als erstes fragen, wie wir die Argumente der zu implementierenden Funktion und die Ergebnisse dieser Funktion in der verwendeten Programmier-Sprache darstellen können. In diesem Fall müssen wir uns also überlegen, wie wir eine aussagenlogische Formel in SETLX repräsentieren können, denn Ergebnisswerte **true** und **false** stehen ja als Wahrheitswerte unmittelbar zur Verfügung. Zusammengesetzte Daten-Strukturen können in SETLX am einfachsten als Terme dargestellt werden und das ist auch der Weg, den wir für die aussagenlogischen Formeln beschreiten werden. Wir definieren die Repräsentation von aussagenlogischen Formeln formal dadurch, dass wir eine Funktion

$$rep : \mathcal{F} \rightarrow \text{SETLX}$$

definieren, die einer aussagenlogischen Formel f einen Term $rep(f)$ zuordnet. Wir werden dabei die in SETLX bereits vorhandenen logischen Operatoren “!”, “&&”, “||”, “=>” und “<==>” benutzen, denn damit können wir die aussagenlogischen Formeln in sehr natürlicher Weise darstellen.

1. \top wird repräsentiert durch den Wahrheitswert **true**.

$$\text{rep}(\top) := \text{true}$$

2. \perp wird repräsentiert durch den Wahrheitswert **false**.

$$\text{rep}(\perp) := \text{false}$$

3. Eine aussagenlogische Variable $p \in \mathcal{P}$ repräsentieren wir durch einen Term der Form

$$\text{\textasciitilde{variable}}(p).$$

Der Grund für diese zunächst seltsam anmutende Darstellung der Variable liegt darin, dass SETLX intern Variablen in der obigen Form darstellt. Wenn wir später einen **String** mit Hilfe der SETLX-Funktion **parse** in eine aussagenlogische Formel umwandeln wollen, so müssen wir Variablen in dieser Form repräsentieren. Damit haben wir also

$$\text{rep}(p) := \text{\textasciitilde{variable}}(p) \quad \text{für alle } p \in \mathcal{P}.$$

4. Ist f eine aussagenlogische Formel, so repräsentieren wir $\neg f$ mit Hilfe des Operators “! ”:

$$\text{rep}(\neg f) := !\text{rep}(f).$$

5. Sind f_1 und f_2 aussagenlogische Formel, so repräsentieren wir $f_1 \vee f_2$ mit Hilfe des Operators “|| ”:

$$\text{rep}(f \vee g) := \text{rep}(f) || \text{rep}(g).$$

6. Sind f_1 und f_2 aussagenlogische Formel, so repräsentieren wir $f_1 \wedge f_2$ mit Hilfe des Operators “&& ”:

$$\text{rep}(f \wedge g) := \text{rep}(f) \&\& \text{rep}(g).$$

7. Sind f_1 und f_2 aussagenlogische Formel, so repräsentieren wir $f_1 \rightarrow f_2$ mit Hilfe des Operators “=> ”:

$$\text{rep}(f \rightarrow g) := \text{rep}(f) \Rightarrow \text{rep}(g).$$

8. Sind f_1 und f_2 aussagenlogische Formel, so repräsentieren wir $f_1 \leftrightarrow f_2$ mit Hilfe des Operators “<==> ”:

$$\text{rep}(f \leftrightarrow g) := \text{rep}(f) \Leftrightarrow \text{rep}(g).$$

Bei der Wahl der Repräsentation, mit der wir eine Formel in SETLX repräsentieren, sind wir weitgehend frei. Wir hätten oben sicher auch eine andere Repräsentation verwenden können. Beispielsweise wurden in einer früheren Version dieses Skriptes die aussagenlogischen Formeln als Listen repräsentiert. Eine gute Repräsentation sollte einerseits möglichst intuitiv sein, andererseits ist es auch wichtig, dass die Repräsentation für die zu entwickelnden Algorithmen adäquat ist. Im wesentlichen heißt dies, dass es einerseits einfach sein sollte, auf die Komponenten einer Formel zuzugreifen, andererseits sollte es auch leicht sein, die entsprechende Repräsentation zu erzeugen. Da wir zur Darstellung der aussagenlogischen Formeln die selben Operatoren verwenden, die auch in SETLX selber benutzt werden, können wir die in SETLX vordefinierte Funktion **parse** benutzen um einen String in eine Formel umzuwandeln. Beispielsweise liefert der Aufruf

```
f := parse("p => p || !q");
```

für f die Formel

```
p => p || !q.
```

Mit Hilfe der SETLX-Funktion **canonical** können wir uns anschauen, wie die Formel in SETLX intern als Term dargestellt wird. Die Eingabe

```
canonical(f);
```

in der Kommandozeile liefert uns das Ergebnis

```
~implication(~variable("p"), ~disjunction(~variable("p"), ~negation(~variable("q"))))
```

Wir erkennen beispielsweise, dass in SETLX der Operator “=>” intern durch das Funktions-Zeichen “~implication” dargestellt wird.

Als nächstes geben wir an, wie wir eine aussagenlogische Interpretation in SETLX darstellen. Eine aussagenlogische Interpretation ist eine Funktion

$$\mathcal{I} : \mathcal{P} \rightarrow \mathbb{B}$$

von der Menge der Aussage-Variablen \mathcal{P} in die Menge der Wahrheitswerte \mathbb{B} . Ist eine Formel f gegeben, so ist klar, dass bei der Interpretation \mathcal{I} nur die Aussage-Variablen p eine Rolle spielen, die auch in der Formel f auftreten. Wir können daher die Interpretation \mathcal{I} durch eine funktionale Relation darstellen, also durch eine Menge von Paaren $[p, b]$, für die p eine Aussage-Variable ist und für die zusätzlich $b \in \mathbb{B}$ gilt:

$$\mathcal{I} \subseteq \mathcal{P} \times \mathbb{B}.$$

Damit können wir jetzt eine einfache Funktion schreiben, dass den Wahrheitswert einer aussagenlogischen Formel f unter einer gegebenen aussagenlogischen Interpretation \mathcal{I} berechnet. Die Funktion `evaluate.stlx` ist in Abbildung 5.1 auf Seite 85 gezeigt.

```

1  evaluate := procedure(f, i) {
2      match (f) {
3          case true:      return true;
4          case false:     return false;
5          case ^variable(p): return i[p];
6          case !g:        return !evaluate(g, i);
7          case g && h:     return evaluate(g, i) && evaluate(h, i);
8          case g || h:     return evaluate(g, i) || evaluate(h, i);
9          case g => h:     return evaluate(g, i) => evaluate(h, i);
10         case g <==> h:   return evaluate(g, i) == evaluate(h, i);
11         default:        abort("syntax error in evaluate($f$, $i$)");
12     }
13 };

```

Abbildung 5.1: Auswertung einer aussagenlogischen Formel.

Wir diskutieren jetzt die Implementierung der Funktion `evaluate()` Zeile für Zeile:

1. Falls die Formel f den Wert `true` hat, so repräsentiert f die Formel \top . Also ist das Ergebnis der Auswertung unabhängig von der aussagenlogischen Interpretation i immer `true`.
2. Falls die Formel f den Wert `false` hat, so repräsentiert f die Formel \perp . Also ist das Ergebnis der Auswertung unabhängig von der aussagenlogischen Interpretation i immer `false`.
3. In Zeile 5 betrachten wir den Fall, dass das Argument f eine aussagenlogische Variable repräsentiert.
In diesem Fall müssen wir die Belegung i , die ja eine Funktion von den aussagenlogischen Variablen in die Wahrheitswerte ist, auf die Variable f anwenden. Da wir die Belegung als eine funktionale Relation dargestellt haben, können wir diese Relation durch den Ausdruck $i[p]$ sehr einfach für die Variable p auswerten.
4. In Zeile 6 betrachten wir den Fall, dass f die Form $!g$ hat und folglich die Formel $\neg g$ repräsentiert. In diesem Fall werten wir erst g unter der Belegung i aus und negieren dann das Ergebnis.
5. In Zeile 7 betrachten wir den Fall, dass f die Form $g_1 \&\& g_2$ hat und folglich die Formel $g_1 \wedge g_2$ repräsentiert. In diesem Fall werten wir zunächst g_1 und g_2 unter der Belegung i aus und verknüpfen das Ergebnis mit dem Operator `"&&"`.
6. In Zeile 8 betrachten wir den Fall, dass f die Form $g_1 || g_2$ hat und folglich die Formel $g_1 \vee g_2$ repräsentiert. In diesem Fall werten wir zunächst g_1 und g_2 unter der Belegung i aus und verknüpfen das Ergebnis mit dem Operator `"||"`.

7. In Zeile 9 betrachten wir den Fall, dass f die Form $g_1 \Rightarrow g_2$ hat und folglich die Formel $g_1 \rightarrow g_2$ repräsentiert. In diesem Fall werten wir zunächst g_1 und g_2 unter der Belegung i aus und benutzen dann den Operator “ \Rightarrow ” der Sprache SETLX.
8. In Zeile 10 führen wir die Auswertung einer Formel $g_1 \Leftrightarrow g_2$ auf die Gleichheit zurück: Die Formel $f \leftrightarrow g$ ist genau dann wahr, wenn f und g den selben Wahrheitswert haben.
9. Wenn keiner der vorhergehenden Fälle greift, liegt ein Syntax-Fehler vor, auf den wir in Zeile 11 hinweisen.

5.3.5 Eine Anwendung

Wir betrachten eine spielerische Anwendung der Aussagenlogik. Inspektor Watson wird zu einem Juweliergeschäft gerufen, in das eingebrochen worden ist. In der unmittelbaren Umgebung werden drei Verdächtige Anton, Bruno und Claus festgenommen. Die Auswertung der Akten ergibt folgendes:

1. Einer der drei Verdächtigen muß die Tat begangen haben:

$$f_1 := a \vee b \vee c.$$

2. Wenn Anton schuldig ist, so hat er genau einen Komplizen.

Diese Aussage zerlegen wir zunächst in zwei Teilaussagen:

- (a) Wenn Anton schuldig ist, dann hat er mindestens einen Komplizen:

$$f_2 := a \rightarrow b \vee c$$

- (b) Wenn Anton schuldig ist, dann hat er höchstens einen Komplizen:

$$f_3 := a \rightarrow \neg(b \wedge c)$$

3. Wenn Bruno unschuldig ist, dann ist auch Claus unschuldig:

$$f_4 := \neg b \rightarrow \neg c$$

4. Wenn genau zwei schuldig sind, dann ist Claus einer von ihnen.

Es ist nicht leicht zu sehen, wie diese Aussage sich aussagenlogisch formulieren läßt. Wir behelfen uns mit einem Trick und überlegen uns, wann die obige Aussage falsch ist. Wir sehen, die Aussage ist dann falsch, wenn Claus nicht schuldig ist und wenn gleichzeitig Anton und Bruno schuldig sind. Damit lautet die Formalisierung der obigen Aussage:

$$f_5 := \neg(\neg c \wedge a \wedge b)$$

5. Wenn Claus unschuldig ist, ist Anton schuldig.

$$f_6 := \neg c \rightarrow a$$

Wir haben nun eine Menge $F = \{f_1, f_2, f_3, f_4, f_5, f_6\}$ von Formeln. Wir fragen uns nun, für welche Belegungen \mathcal{I} alle Formeln aus F wahr werden. Wenn es genau eine Belegungen gibt, für die dies der Fall ist, dann liefert uns die Belegung den oder die Täter. Eine Belegung entspricht dabei 1-zu-1 der Menge der Täter. Hätten wir beispielsweise

$$\mathcal{I} = \{\langle a, \text{false} \rangle, \langle b, \text{false} \rangle, \langle c, \text{true} \rangle\}.$$

so wäre Claus der alleinige Täter. Diese Belegung löst unser Problem offenbar nicht, denn Sie widerspricht der dritten Aussage: Da Bruno unschuldig wäre, wäre dann auch Claus unschuldig. Da es zu zeitraubend ist, alle Belegungen von Hand auszuprobieren, schreiben wir besser ein Programm, das für uns die notwendige Berechnung durchführt. Abbildung 5.2 zeigt das Programm `watson.stlx`. Wir diskutieren diese Programm nun Zeile für Zeile.

1. In den Zeilen 6 – 16 definieren wir die Formeln f_1, \dots, f_6 . Wir müssen hier die Formeln in die SETLX-Repräsentation bringen. Diese Arbeit wird uns durch die Benutzung der Funktion `parse` leicht gemacht.

```

1  // This procedure turns a subset m of set set of all variables
2  // into a propositional valuation i, such that i(x) is true
3  // iff x is an element of m.
4  createValuation := procedure(m, v) {
5      return { [ x, x in m ] : x in v };
6  };
7  // Austin, Brian, or Colin is guilty.
8  f1 := parse("a || b || c");
9  // If Austin is guilty, he has exactly one accomplice.
10 f2 := parse("a => b || c");    // at least one accomplice
11 f3 := parse("a => !(b && c)"); // at most one accomplice
12 // If Brian is innocent, then Colin is innocent, too.
13 f4 := parse("!b => !c");
14 // If exactly two are guilty, then Colin is one of them.
15 f5 := parse("!(a && b && !c)");
16 // If Colin is innocent, then Austin is guilty.
17 f6 := parse("!c => a");
18 fs := { f1, f2, f3, f4, f5, f6 };
19 v := { "a", "b", "c" };
20 p := 2 ** v;
21 print("p = ", p);
22 // b is the set of all propositional valuations.
23 b := { createValuation(m, v) : m in p };
24 s := { i in b | forall (f in fs | evaluate(f, i)) };
25 print("Set of all valuations satisfying all facts: ", s);
26 if (#s == 1) {
27     i := arb(s);
28     taeter := { x in v | i[x] };
29     print("Set of culprits: ", taeter);
30 }

```

Abbildung 5.2: Programm zur Aufklärung des Einbruchs.

2. Als nächstes müssen wir uns überlegen, wie wir alle Belegungen aufzählen können. Wir hatten oben schon beobachtet, dass die Belegungen 1-zu-1 zu den möglichen Mengen der Täter korrespondieren. Die Mengen der möglichen Täter sind aber alle Teilmengen der Menge

$$\{\text{"a"}, \text{"b"}, \text{"c"}\}.$$

Wir berechnen daher in Zeile 18 zunächst die Menge aller dieser Teilmengen.

3. Wir brauchen jetzt eine Möglichkeit, eine Teilmenge in eine Belegung umzuformen. In den Zeilen 3 – 5 haben wir eine Prozedur implementiert, die genau dies leistet. Um zu verstehen, wie diese Funktion arbeitet, betrachten wir ein Beispiel und nehmen an, dass wir aus der Menge

$$m = \{\text{"a"}, \text{"c"}\}$$

eine Belegung \mathcal{I} erstellen sollen. Wir erhalten dann

$$\mathcal{I} = \{\langle \text{"a"}, \text{true} \rangle, \langle \text{"b"}, \text{false} \rangle, \langle \text{"c"}, \text{true} \rangle\}.$$

Das allgemeine Prinzip ist offenbar, dass für eine aussagenlogische Variable x das Paar $\langle x, \text{true} \rangle$ genau dann in der Belegung \mathcal{I} enthalten ist, wenn $x \in m$ ist, andernfalls ist das Paar $\langle x, \text{false} \rangle$ in \mathcal{I} . Damit könnten wir die Menge aller Belegungen, die genau die Elemente aus m wahrmachen, wie folgt schreiben:

$$\{ [x, \text{true}] : x \in m \} + \{ [x, \text{false}] : x \in A \mid \neg(x \in m) \}$$

Es geht aber einfacher, denn wir können beide Fälle zusammenfassen, indem wir fordern, dass das Paar $\langle x, x \in m \rangle$ ein Element der Belegung \mathcal{I} ist. Genau das steht in Zeile 4.

4. In Zeile 22 sammeln wir in der Menge b alle möglichen Belegungen auf.
5. In Zeile 23 berechnen wir die Menge s aller der Belegungen i , für die alle Formeln aus der Menge fs wahr werden.
6. Falls es genau eine Belegung gibt, die alle Formeln wahr macht, dann haben wir das Problem lösen können. In diesem Fall extrahieren wir in Zeile 26 diese Belegungen aus der Menge s und geben anschließend die Menge der Täter aus.

Lassen wir das Programm laufen, so erhalten wir als Ausgabe

Set of culprits: {"b", "c"}

Damit liefern unsere ursprünglichen Formeln ausreichende Information um die Täter zu überführen: Bruno und Claus sind schuldig.

5.4 Tautologien

Die Tabelle in Abbildung 5.2 zeigt, dass die Formel

$$(p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q$$

für jede aussagenlogische Interpretation wahr ist, denn in der letzten Spalte dieser Tabelle steht immer der Wert **true**. Formeln mit dieser Eigenschaft bezeichnen wir als *Tautologie*.

Definition 22 (Tautologie) Ist f eine aussagenlogische Formel und gilt

$$\mathcal{I}(f) = \text{true} \quad \text{für jede aussagenlogische Interpretation } \mathcal{I},$$

dann ist f eine *Tautologie*. In diesem Fall schreiben wir

$$\models f.$$

□

Ist eine Formel f eine Tautologie, so sagen wir auch, dass f *allgemeingültig* ist.

Beispiele:

1. $\models p \vee \neg p$
2. $\models p \rightarrow p$
3. $\models p \wedge q \rightarrow p$
4. $\models p \rightarrow p \vee q$
5. $\models (p \rightarrow \perp) \leftrightarrow \neg p$
6. $\models p \wedge q \leftrightarrow q \wedge p$

Wir können die Tatsache, dass es sich bei diesen Formeln um Tautologien handelt, durch eine Tabelle nachweisen, die analog zu der auf Seite 82 gezeigten Tabelle 5.2 aufgebaut ist. Dieses Verfahren ist zwar konzeptuell sehr einfach, allerdings zu ineffizient, wenn die Anzahl der aussagenlogischen Variablen groß ist. Ziel dieses Kapitels ist daher die Entwicklung besserer Verfahren.

Die letzten beiden Beispiele in der obigen Aufzählung geben Anlaß zu einer neuen Definition.

Definition 23 (Äquivalent) Zwei Formeln f und g heißen *äquivalent* g.d.w.

$$\models f \leftrightarrow g$$

gilt.

□

Beispiele: Es gelten die folgenden Äquivalenzen:

$\models \neg \perp \leftrightarrow \top$	$\models \neg \top \leftrightarrow \perp$	
$\models p \vee \neg p \leftrightarrow \top$	$\models p \wedge \neg p \leftrightarrow \perp$	Tertium-non-Datur
$\models p \vee \perp \leftrightarrow p$	$\models p \wedge \top \leftrightarrow p$	Neutrales Element
$\models p \vee \top \leftrightarrow \top$	$\models p \wedge \perp \leftrightarrow \perp$	
$\models p \wedge p \leftrightarrow p$	$\models p \vee p \leftrightarrow p$	Idempotenz
$\models p \wedge q \leftrightarrow q \wedge p$	$\models p \vee q \leftrightarrow q \vee p$	Kommutativität
$\models (p \wedge q) \wedge r \leftrightarrow p \wedge (q \wedge r)$	$\models (p \vee q) \vee r \leftrightarrow p \vee (q \vee r)$	Assoziativität
$\models \neg \neg p \leftrightarrow p$		Elimination von $\neg \neg$
$\models p \wedge (p \vee q) \leftrightarrow p$	$\models p \vee (p \wedge q) \leftrightarrow p$	Absorption
$\models p \wedge (q \vee r) \leftrightarrow (p \wedge q) \vee (p \wedge r)$	$\models p \vee (q \wedge r) \leftrightarrow (p \vee q) \wedge (p \vee r)$	Distributivität
$\models \neg(p \wedge q) \leftrightarrow \neg p \vee \neg q$	$\models \neg(p \vee q) \leftrightarrow \neg p \wedge \neg q$	DeMorgan'sche Regeln
$\models (p \rightarrow q) \leftrightarrow \neg p \vee q$		Elimination von \rightarrow
$\models (p \leftrightarrow q) \leftrightarrow (\neg p \vee q) \wedge (\neg q \vee p)$		Elimination von \leftrightarrow

Wir können diese Äquivalenzen nachweisen, indem wir in einer Tabelle sämtliche Belegungen durchprobieren. Eine solche Tabelle heißt auch *Wahrheits-Tafel*. Wir demonstrieren dieses Verfahren anhand der ersten DeMorgan'schen Regel. Wir erkennen, dass in Abbildung 5.3 in den letzten beiden Spalten in jeder Zeile die selben

p	q	$\neg p$	$\neg q$	$p \wedge q$	$\neg(p \wedge q)$	$\neg p \vee \neg q$
true	true	false	false	true	false	false
true	false	false	true	false	true	true
false	true	true	false	false	true	true
false	false	true	true	false	true	true

Tabelle 5.3: Nachweis der ersten DeMorgan'schen Regel.

Werte stehen. Daher sind die Formeln, die zu diesen Spalten gehören, äquivalent.

5.4.1 Testen der Allgemeingültigkeit in SetLX

Die manuelle Überprüfung der Frage, ob eine gegebene Formel f eine Tautologie ist, läuft auf die Erstellung umfangreicher Wahrheitstabellen heraus. Solche Wahrheitstabellen von Hand zu erstellen ist viel zu zeitaufwendig. Wir wollen daher nun ein SETLX-Programm entwickeln, mit dessen Hilfe wir die obige Frage automatisch beantworten können. Die Grundidee ist, dass wir die zu untersuchende Formel für alle möglichen Belegungen auswerten und überprüfen, dass sich bei der Auswertung jedesmal der Wert **true** ergibt. Dazu müssen wir zunächst einen Weg finden, alle möglichen Belegungen einer Formel zu berechnen. Wir haben früher schon gesehen, dass Belegungen \mathcal{I} zu Teilmengen M der Menge der aussagenlogischen Variablen \mathcal{P} korrespondieren, denn für jedes $M \subseteq \mathcal{P}$ können wir eine aussagenlogische Belegung $\mathcal{I}(M)$ wie folgt definieren:

$$\mathcal{I}(M)(p) := \begin{cases} \text{true} & \text{falls } p \in M; \\ \text{false} & \text{falls } p \notin M. \end{cases}$$

Um die aussagenlogische Belegung \mathcal{I} in SETLX darstellen zu können, fassen wir die Belegung \mathcal{I} als links-totale und rechts-eindeutige Relation $\mathcal{I} \subseteq \mathcal{P} \times \mathbb{B}$ auf. Dann haben wir

$$\mathcal{I} = \{ \langle p, \text{true} \rangle \mid p \in M \} \cup \{ \langle p, \text{false} \rangle \mid p \notin M \}.$$

Dies lässt sich noch zu

$$\mathcal{I} = \{ \langle p, p \in M \rangle \mid p \in \mathcal{P} \}$$

vereinfachen. Mit dieser Idee können wir nun eine Prozedur implementieren, die für eine gegebene aussagenlogische Formel f testet, ob f eine Tautologie ist.

```

1  tautology := procedure(f) {
2      p := collectVars(f);
3      // a is the set of all propositional valuations.
4      a := { { [x, x in m] : x in p } : m in pow(p) };
5      if (forall (i in a | evaluate(f, i))) {
6          return {};
7      } else {
8          return arb({ i in a | !evaluate(f, i) });
9      }
10 };
11 collectVars := procedure(f) {
12     match (f) {
13         case true:      return {};
14         case false:     return {};
15         case ^variable(p): return { p };
16         case !g:        return collectVars(g);
17         case g && h:     return collectVars(g) + collectVars(h);
18         case g || h:     return collectVars(g) + collectVars(h);
19         case g => h:     return collectVars(g) + collectVars(h);
20         case g <==> h:   return collectVars(g) + collectVars(h);
21         default:        abort("syntax error in collectVars($f$)");
22     }
23 };

```

Abbildung 5.3: Überprüfung der Allgemeingültigkeit einer aussagenlogischen Formel.

Die in Abbildung 5.3 auf Seite 90 gezeigte Funktion **tautology** testet, ob die als Argument übergebene aussagenlogische Formel f allgemeingültig ist. Die Prozedur verwendet die Funktion **evaluate** aus dem in Abbildung 5.1 auf Seite 85 gezeigten Programm. Wir diskutieren die Definition der Funktion **tautology** nun Zeile für Zeile:

1. In Zeile 2 sammeln wir alle aussagenlogischen Variablen auf, die in der zu überprüfenden Formel auftreten. Die dazu benötigte Prozedur **collectVars** ist in den Zeilen 11 – 23 gezeigt. Diese Prozedur ist durch Induktion über den Aufbau einer Formel definiert und liefert als Ergebnis die Menge aller Aussage-Variablen, die in der aussagenlogischen Formel f auftreten.

Es ist klar, dass bei der Berechnung von $\mathcal{I}(f)$ für eine Formel f und eine aussagenlogische Interpretation \mathcal{I} nur die Werte von $\mathcal{I}(p)$ eine Rolle spielen, für die die Variable p in f auftritt. Zur Analyse von f können wir uns also auf aussagenlogische Interpretationen der Form

$$\mathcal{I} : \mathcal{P} \rightarrow \mathbb{B} \quad \text{mit} \quad \mathcal{P} = \text{collectVars}(f)$$

beschränken.

2. In Zeile 4 berechnen wir die Menge aller aussagenlogischen Interpretationen über der Menge \mathcal{P} der aussagenlogischen Variablen. Wir berechnen für eine Menge m von aussagenlogischen Variablen die Interpretation $\mathcal{I}(m)$ wie oben diskutiert mit Hilfe der Formel

$$\mathcal{I}(m) := \{ \langle x, x \in m \rangle \mid x \in V \}.$$

Betrachten wir zur Verdeutlichung als Beispiel die Formel

$$\neg(p \wedge q) \leftrightarrow \neg p \vee \neg q.$$

Die Menge \mathcal{P} der aussagenlogischen Variablen, die in dieser Formel auftreten, ist

$$\mathcal{P} = \{p, q\}.$$

Die Potenz-Menge der Menge \mathcal{P} ist

$$2^{\mathcal{P}} = \{\{\}, \{p\}, \{q\}, \{p, q\}\}.$$

Wir bezeichnen die vier Elemente dieser Menge mit m_1, m_2, m_3, m_4 :

$$m_1 := \{\}, m_2 := \{p\}, m_3 := \{q\}, m_4 := \{p, q\}.$$

Aus jeder dieser Mengen m_i gewinnen wir nun eine aussagenlogische Interpretation $\mathcal{I}(m_i)$:

$$\mathcal{I}(m_1) := \{\langle x, x \in \{\} \rangle \mid x \in \{p, q\}\} = \{\langle p, p \in \{\} \rangle, \langle q, q \in \{\} \rangle\} = \{\langle p, \text{false} \rangle, \langle q, \text{false} \rangle\}.$$

$$\mathcal{I}(m_2) := \{\langle x, x \in \{p\} \rangle \mid x \in \{p, q\}\} = \{\langle p, p \in \{p\} \rangle, \langle q, q \in \{p\} \rangle\} = \{\langle p, \text{true} \rangle, \langle q, \text{false} \rangle\}.$$

$$\mathcal{I}(m_3) := \{\langle x, x \in \{q\} \rangle \mid x \in \{p, q\}\} = \{\langle p, p \in \{q\} \rangle, \langle q, q \in \{q\} \rangle\} = \{\langle p, \text{false} \rangle, \langle q, \text{true} \rangle\}.$$

$$\mathcal{I}(m_4) := \{\langle x, x \in \{p, q\} \rangle \mid x \in \{p, q\}\} = \{\langle p, p \in \{p, q\} \rangle, \langle q, q \in \{p, q\} \rangle\} = \{\langle p, \text{true} \rangle, \langle q, \text{true} \rangle\}.$$

damit haben wir alle möglichen Interpretationen der Variablen p und q .

3. In Zeile 5 testen wir, ob die Formel f für alle möglichen Interpretationen i aus der Menge a aller Interpretationen wahr ist. Ist dies der Fall, so geben wir die leere Menge als Ergebnis zurück.

Falls es allerdings eine Belegungen i in der Menge a gibt, für die die Auswertung von f den Wert *false* liefert, so bilden wir in Zeile 8 die Menge aller solcher Belegungen und wählen mit Hilfe der Funktion *arb* eine beliebige Belegungen aus dieser Menge aus, die wir dann als Gegenbeispiel zurück geben.

5.4.2 Nachweis der Allgemeingültigkeit durch Äquivalenz-Umformungen

Wollen wir nachweisen, dass eine Formel eine Tautologie ist, können wir uns prinzipiell immer einer Wahrheits-Tafel bedienen. Aber diese Methode hat einen Haken: Kommen in der Formel n verschiedene Aussage-Variablen vor, so hat die Tabelle 2^n Zeilen. Beispielsweise hat die Tabelle zum Nachweis eines der Distributiv-Gesetze bereits 8 Zeilen, da hier 3 verschiedene Variablen auftreten. Eine andere Möglichkeit nachzuweisen, dass eine Formel eine Tautologie ist, ergibt sich dadurch, dass wir die Formel mit Hilfe der oben aufgeführten Äquivalenzen *vereinfachen*. Wenn es gelingt, eine Formel F unter Verwendung dieser Äquivalenzen zu \top zu vereinfachen, dann ist gezeigt, dass F eine Tautologie ist. Wir demonstrieren das Verfahren zunächst an einem Beispiel. Mit Hilfe einer Wahrheits-Tafel hatten wir schon gezeigt, dass die Formel

$$(p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q$$

eine Tautologie ist. Wir zeigen nun, wie wir diesen Tatbestand auch durch eine Kette von Äquivalenz-Umformungen einsehen können:

	$(p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q$	(Elimination von \rightarrow)
\Leftrightarrow	$(\neg p \vee q) \rightarrow (\neg p \rightarrow q) \rightarrow q$	(Elimination von \rightarrow)
\Leftrightarrow	$(\neg p \vee q) \rightarrow (\neg \neg p \vee q) \rightarrow q$	(Elimination der Doppelnegation)
\Leftrightarrow	$(\neg p \vee q) \rightarrow (p \vee q) \rightarrow q$	(Elimination von \rightarrow)
\Leftrightarrow	$\neg(\neg p \vee q) \vee ((p \vee q) \rightarrow q)$	(DeMorgan)
\Leftrightarrow	$(\neg \neg p \wedge \neg q) \vee ((p \vee q) \rightarrow q)$	(Elimination der Doppelnegation)
\Leftrightarrow	$(p \wedge \neg q) \vee ((p \vee q) \rightarrow q)$	(Elimination von \rightarrow)
\Leftrightarrow	$(p \wedge \neg q) \vee (\neg(p \vee q) \vee q)$	(DeMorgan)
\Leftrightarrow	$(p \wedge \neg q) \vee ((\neg p \wedge \neg q) \vee q)$	(Distributivität)
\Leftrightarrow	$(p \wedge \neg q) \vee ((\neg p \vee q) \wedge (\neg q \vee q))$	(Tertium-non-Datur)
\Leftrightarrow	$(p \wedge \neg q) \vee ((\neg p \vee q) \wedge \top)$	(Neutrales Element)
\Leftrightarrow	$(p \wedge \neg q) \vee (\neg p \vee q)$	(Distributivität)
\Leftrightarrow	$(p \vee (\neg p \vee q)) \wedge (\neg q \vee (\neg p \vee q))$	(Assoziativität)
\Leftrightarrow	$((p \vee \neg p) \vee q) \wedge (\neg q \vee (\neg p \vee q))$	(Tertium-non-Datur)
\Leftrightarrow	$(\top \vee q) \wedge (\neg q \vee (\neg p \vee q))$	(Neutrales Element)
\Leftrightarrow	$\top \wedge (\neg q \vee (\neg p \vee q))$	(Neutrales Element)
\Leftrightarrow	$\neg q \vee (\neg p \vee q)$	(Assoziativität)
\Leftrightarrow	$(\neg q \vee \neg p) \vee q$	(Kommutativität)
\Leftrightarrow	$(\neg p \vee \neg q) \vee q$	(Assoziativität)
\Leftrightarrow	$\neg p \vee (\neg q \vee q)$	(Tertium-non-Datur)
\Leftrightarrow	$\neg p \vee \top$	(Neutrales Element)
\Leftrightarrow	\top	

Die Umformungen in dem obigen Beweis sind nach einem bestimmten System durchgeführt worden. Um dieses System präzise formulieren zu können, benötigen wir noch einige Definitionen.

Definition 24 (Literal) Eine aussagenlogische Formel f heißt *Literal* g.d.w. einer der folgenden Fälle vorliegt:

1. $f = \top$ oder $f = \perp$.
2. $f = p$, wobei p eine aussagenlogische Variable ist.
In diesem Fall sprechen wir von einem *positiven* Literal.
3. $f = \neg p$, wobei p eine aussagenlogische Variable ist.
In diesem Fall sprechen wir von einem *negativen* Literal.

Die Menge aller Literale bezeichnen wir mit \mathcal{L} . □

Später werden wir noch den Begriff des *Komplements* eines Literals benötigen. Ist l ein Literal, so wird das Komplement von l mit \bar{l} bezeichnet. Das Komplement wird durch Fall-Unterscheidung definiert:

1. $\bar{\top} = \perp$ und $\bar{\perp} = \top$.
2. $\bar{p} := \neg p$, falls $p \in \mathcal{P}$.
3. $\bar{\neg p} := p$, falls $p \in \mathcal{P}$.

Wir sehen, dass das Komplement \bar{l} eines Literals l äquivalent zur Negation von l ist, wir haben also

$$\models \bar{l} \leftrightarrow \neg l.$$

Definition 25 (Klausel) Eine aussagenlogische Formel k ist eine *Klausel* wenn k die Form

$$k = l_1 \vee \dots \vee l_r$$

hat, wobei l_i für alle $i = 1, \dots, r$ ein Literal ist. Eine Klausel ist also eine Disjunktion von Literalen. Die Menge aller Klauseln bezeichnen wir mit \mathcal{K} . □

Oft werden Klauseln auch einfach als *Mengen* von Literalen betrachtet. Durch diese Sichtweise abstrahieren wir von der Reihenfolge und der Anzahl des Auftretens der Literale in der Disjunktion. Dies ist möglich aufgrund der Assoziativität, Kommutativität und Idempotenz des Junktors “ \vee ”. Für die Klausel $l_1 \vee \dots \vee l_r$ schreiben wir also in Zukunft auch

$$\{l_1, \dots, l_r\}.$$

Das folgende Beispiel illustriert die Nützlichkeit der Mengen-Schreibweise von Klauseln. Wir betrachten die beiden Klauseln

$$p \vee q \vee \neg r \vee p \quad \text{und} \quad \neg r \vee q \vee \neg r \vee p.$$

Die beiden Klauseln sind zwar äquivalent, aber die Formeln sind verschieden. Überführen wir die beiden Klauseln in Mengen-Schreibweise, so erhalten wir

$$\{p, q, \neg r\} \quad \text{und} \quad \{\neg r, q, p\}.$$

In einer Menge kommt jedes Element höchstens einmal vor und die Reihenfolge, in der die Elemente auftreten, spielt auch keine Rolle. Daher sind die beiden obigen Mengen gleich! Durch die Tatsache, dass Mengen von der Reihenfolge und der Anzahl der Elemente abstrahieren, implementiert die Mengen-Schreibweise die Assoziativität, Kommutativität und Idempotenz der Disjunktion. Übertragen wir die aussagenlogische Äquivalenz

$$l_1 \vee \dots \vee l_r \vee \perp \leftrightarrow l_1 \vee \dots \vee l_r$$

in Mengen-Schreibweise, so erhalten wir

$$\{l_1, \dots, l_r, \perp\} \leftrightarrow \{l_1, \dots, l_r\}.$$

Dies zeigt, dass wir das Element \perp in einer Klausel getrost weglassen können. Betrachten wir die letzten Äquivalenz für den Fall, dass $r = 0$ ist, so haben wir

$$\{\perp\} \leftrightarrow \{\}.$$

Damit sehen wir, dass die leere Menge von Literalen als \perp zu interpretieren ist.

Definition 26 Eine Klausel k ist *trivial*, wenn einer der beiden folgenden Fälle vorliegt:

1. $\top \in k$.
2. Es existiert $p \in \mathcal{P}$ mit $p \in k$ und $\neg p \in k$.

In diesem Fall bezeichnen wir p und $\neg p$ als *komplementäre Literale*. □

Satz 27 Eine Klausel ist genau dann eine Tautologie, wenn sie trivial ist.

Beweis: Wir nehmen zunächst an, dass die Klausel k trivial ist. Falls nun $\top \in k$ ist, dann gilt wegen der Gültigkeit der Äquivalenz $f \vee \top \leftrightarrow \top$ offenbar $k \leftrightarrow \top$. Ist p eine Aussage-Variable, so dass sowohl $p \in k$ als auch $\neg p \in k$ gilt, dann folgt aufgrund der Äquivalenz $p \vee \neg p \leftrightarrow \top$ sofort $k \leftrightarrow \top$.

Wir nehmen nun an, dass die Klausel k eine Tautologie ist. Wir führen den Beweis indirekt und nehmen an, dass k nicht trivial ist. Damit gilt $\top \notin k$ und k kann auch keine komplementären Literale enthalten. Damit hat k dann die Form

$$k = \{\neg p_1, \dots, \neg p_m, q_1, \dots, q_n\} \quad \text{mit } p_i \neq q_j \text{ für alle } i \in \{1, \dots, m\} \text{ und } j \in \{1, \dots, n\}.$$

Dann könnten wir eine Interpretation \mathcal{I} wie folgt definieren:

1. $\mathcal{I}(p_i) = \mathbf{true}$ für alle $i = 1, \dots, m$ und
2. $\mathcal{I}(q_j) = \mathbf{false}$ für alle $j = 1, \dots, n$,

Mit dieser Interpretation würde offenbar $\mathcal{I}(k) = \mathbf{false}$ gelten und damit könnte k keine Tautologie sein. Also ist die Annahme, dass k nicht trivial ist, falsch. □

Definition 28 (Konjunktive Normalform) Eine Formel f ist in *konjunktiver Normalform* (kurz KNF) genau dann, wenn f eine Konjunktion von Klauseln ist, wenn also gilt

$$f = k_1 \wedge \cdots \wedge k_n,$$

wobei die k_i für alle $i = 1, \dots, n$ Klauseln sind. □

Aus der Definition der KNF folgt sofort:

Korollar 29 Ist $f = k_1 \wedge \cdots \wedge k_n$ in konjunktiver Normalform, so gilt

$$\models f \quad \text{genau dann, wenn} \quad \models k_i \quad \text{für alle } i = 1, \dots, n. \quad \square$$

Damit können wir für eine Formel $f = k_1 \wedge \cdots \wedge k_n$ in konjunktiver Normalform leicht entscheiden, ob f eine Tautologie ist, denn f ist genau dann eine Tautologie, wenn alle Klauseln k_i trivial sind.

Da für die Konjunktion genau wie für die Disjunktion Assoziativ-Gesetz, Kommutativ-Gesetz und Idempotenz-Gesetz gilt, ist es zweckmäßig, auch für Formeln in konjunktiver Normalform eine Mengen-Schreibweise einzuführen. Ist also die Formel

$$f = k_1 \wedge \cdots \wedge k_n$$

in konjunktiver Normalform, so repräsentieren wir diese Formel durch die Menge ihrer Klauseln und schreiben

$$f = \{k_1, \dots, k_n\}.$$

Wir geben ein Beispiel: Sind p, q und r Aussage-Variablen, so ist die Formel

$$(p \vee q \vee \neg r) \wedge (q \vee \neg r \vee p \vee q) \wedge (\neg r \vee p \vee \neg q)$$

in konjunktiver Normalform. In Mengen-Schreibweise wird daraus

$$\{\{p, q, \neg r\}, \{p, \neg q, \neg r\}\}.$$

Wir stellen nun ein Verfahren vor, mit dem sich jede Formel in KNF transformieren läßt. Nach dem oben Gesagten können wir dann leicht entscheiden, ob f eine Tautologie ist.

1. Eliminiere alle Vorkommen des Junktors “ \leftrightarrow ” mit Hilfe der Äquivalenz

$$(f \leftrightarrow g) \leftrightarrow (f \rightarrow g) \wedge (g \rightarrow f)$$

2. Eliminiere alle Vorkommen des Junktors “ \rightarrow ” mit Hilfe der Äquivalenz

$$(f \rightarrow g) \leftrightarrow \neg f \vee g$$

3. Schiebe die Negationszeichen soweit es geht nach innen. Verwende dazu die folgenden Äquivalenzen:

$$(a) \quad \neg \perp \leftrightarrow \top$$

$$(b) \quad \neg \top \leftrightarrow \perp$$

$$(c) \quad \neg \neg f \leftrightarrow f$$

$$(d) \quad \neg(f \wedge g) \leftrightarrow \neg f \vee \neg g$$

$$(e) \quad \neg(f \vee g) \leftrightarrow \neg f \wedge \neg g$$

In dem Ergebnis, das wir nach diesem Schritt erhalten, stehen die Negationszeichen nur noch unmittelbar vor den aussagenlogischen Variablen. Formeln mit dieser Eigenschaft bezeichnen wir auch als Formeln in *Negations-Normalform*.

4. Stehen in der Formel jetzt “ \vee ”-Junktoren über “ \wedge ”-Junktoren, so können wir durch *Ausmultiplizieren*, sprich Verwendung der Distributiv-Gesetze

$$f \vee (g \wedge h) \leftrightarrow (f \vee g) \wedge (f \vee h) \quad \text{und} \quad (f \wedge g) \vee h \leftrightarrow (f \vee h) \wedge (g \vee h)$$

diese Junktoren nach innen schieben.

5. In einem letzten Schritt überführen wir die Formel nun in Mengen-Schreibweise, indem wir zunächst die Disjunktionen aller Literale als Mengen zusammenfassen und anschließend alle so entstandenen Klauseln wieder in einer Menge zusammen fassen.

Hier sollten wir noch bemerken, dass die Formel beim Ausmultiplizieren stark anwachsen kann. Das liegt daran, dass die Formel f auf der rechten Seite der Äquivalenz $f \vee (g \wedge h) \leftrightarrow (f \vee g) \wedge (f \vee h)$ zweimal auftritt, während sie links nur einmal vorkommt.

Wir demonstrieren das Verfahren am Beispiel der Formel

$$(p \rightarrow q) \rightarrow (\neg p \rightarrow \neg q).$$

1. Da die Formel den Junktor “ \leftrightarrow ” nicht enthält, ist im ersten Schritt nichts zu tun.
2. Die Elimination von “ \rightarrow ” liefert

$$\neg(\neg p \vee q) \vee (\neg\neg p \vee \neg q).$$

3. Die Umrechnung auf Negations-Normalform liefert

$$(p \wedge \neg q) \vee (p \vee \neg q).$$

4. Durch “Ausmultiplizieren” erhalten wir

$$(p \vee (p \vee \neg q)) \wedge (\neg q \vee (p \vee \neg q)).$$

5. Die Überführung in die Mengen-Schreibweise ergibt zunächst als Klauseln die beiden Mengen

$$\{p, p, \neg q\} \quad \text{und} \quad \{\neg q, p, \neg q\}.$$

Da die Reihenfolge der Elemente einer Menge aber unwichtig ist und außerdem eine Menge jedes Element nur einmal enthält, stellen wir fest, dass diese beiden Klauseln gleich sind. Fassen wir jetzt die Klauseln noch in einer Menge zusammen, so erhalten wir

$$\{\{p, \neg q\}\}.$$

Beachten Sie, dass sich die Formel durch die Überführung in Mengen-Schreibweise noch einmal deutlich vereinfacht hat.

Damit ist die Formel in KNF überführt.

5.4.3 Berechnung der konjunktiven Normalform in SetIX

Wir geben nun eine Reihe von Prozeduren an, mit deren Hilfe sich eine gegebene Formel f in konjunktive Normalform überführen läßt. Wir beginnen mit einer Prozedur

$$\text{elimGdw} : \mathcal{F} \rightarrow \mathcal{F}$$

die die Aufgabe hat, eine vorgegebene aussagenlogische Formel f in eine äquivalente Formel umzuformen, die den Junktor “ \leftrightarrow ” nicht mehr enthält. Die Funktion $\text{elimGdw}(f)$ wird durch Induktion über den Aufbau der aussagenlogischen Formel f definiert. Dazu stellen wir zunächst rekursive Gleichungen auf, die das Verhalten der Funktion $\text{elimGdw}()$ beschreiben:

1. Wenn f eine Aussage-Variable p ist, so ist nichts zu tun:

$$\text{elimGdw}(p) = p \quad \text{für alle } p \in \mathcal{P}.$$

2. Hat f die Form $f = \neg g$, so eliminieren wir den Junktor “ \leftrightarrow ” aus der Formel g :

$$\text{elimGdw}(\neg g) = \neg \text{elimGdw}(g).$$

3. Im Falle $f = g_1 \wedge g_2$ eliminieren wir den Junktor “ \leftrightarrow ” aus den Formeln g_1 und g_2 :

$$\text{elimGdw}(g_1 \wedge g_2) = \text{elimGdw}(g_1) \wedge \text{elimGdw}(g_2).$$

4. Im Falle $f = g_1 \vee g_2$ eliminieren wir den Junktor “ \leftrightarrow ” aus den Formeln g_1 und g_2 :

$$\text{elimGdw}(g_1 \vee g_2) = \text{elimGdw}(g_1) \vee \text{elimGdw}(g_2).$$

5. Im Falle $f = g_1 \rightarrow g_2$ eliminieren wir den Junktor “ \leftrightarrow ” aus den Formeln g_1 und g_2 :

$$\text{elimGdw}(g_1 \rightarrow g_2) = \text{elimGdw}(g_1) \rightarrow \text{elimGdw}(g_2).$$

6. Hat f die Form $f = g_1 \leftrightarrow g_2$, so benutzen wir die Äquivalenz

$$(g_1 \leftrightarrow g_2) \leftrightarrow ((g_1 \rightarrow g_2) \wedge (g_2 \rightarrow g_1)).$$

Das führt auf die Gleichung:

$$\text{elimGdw}(g_1 \leftrightarrow g_2) = \text{elimGdw}((g_1 \rightarrow g_2) \wedge (g_2 \rightarrow g_1)).$$

Der Aufruf von `elimGdw` auf der rechten Seite der Gleichung ist notwendig, denn der Junktor “ \leftrightarrow ” kann ja noch in g_1 und g_2 auftreten.

Abbildung 5.4 auf Seite 96 zeigt die Implementierung der Prozedur `elimGdw`.

```

1  elimGdw := procedure(f) {
2      match (f) {
3          case !g      : return !elimGdw(g);
4          case g && h   : return elimGdw(g) && elimGdw(h);
5          case g || h   : return elimGdw(g) || elimGdw(h);
6          case g => h   : return elimGdw(g) => elimGdw(h);
7          case g <==> h : return (elimGdw(g) => elimGdw(h)) && (elimGdw(h) => elimGdw(g));
8          default      : return f;
9      }
10 };

```

Abbildung 5.4: Elimination von \leftrightarrow .

Als nächstes betrachten wir die Prozedur zur Elimination des Junktors “ \rightarrow ”. Abbildung 5.5 auf Seite 96 zeigt die Implementierung der Funktion `elimFolgt`. Die der Implementierung zu Grunde liegende Idee ist die selbe wie bei der Elimination des Junktors “ \leftrightarrow ”. Der einzige Unterschied besteht darin, dass wir jetzt die Äquivalenz

$$(g_1 \rightarrow g_2) \leftrightarrow (\neg g_1 \vee g_2)$$

benutzen. Außerdem können wir schon voraussetzen, dass der Junktor “ \leftrightarrow ” bereits vorher eliminiert wurde. Dadurch entfällt ein Fall.

```

1  elimFolgt := procedure(f) {
2      match (f) {
3          case !g      : return !elimFolgt(g);
4          case g && h   : return elimFolgt(g) && elimFolgt(h);
5          case g || h   : return elimFolgt(g) || elimFolgt(h);
6          case g => h   : return !elimFolgt(g) || elimFolgt(h);
7          default      : return f;
8      }
9  };

```

Abbildung 5.5: Elimination von \rightarrow .

Als nächstes zeigen wir die Routinen zur Berechnung der Negations-Normalform. Abbildung 5.6 auf Seite 98 zeigt die Implementierung der Funktionen `nnf` und `neg`, die sich wechselseitig aufrufen. Dabei berechnet `neg(f)`

die Negations-Normalform von $\neg f$, während $\mathbf{nnf}(f)$ die Negations-Normalform von f berechnet, es gilt also

$$\mathbf{neg}(f) = \mathbf{nnf}(\neg f).$$

Die eigentliche Arbeit wird dabei in der Funktion **neg** erledigt, denn dort kommen die beiden DeMorgan'schen Gesetze

$$\neg(f \wedge g) \leftrightarrow (\neg f \vee \neg g) \quad \text{und} \quad \neg(f \vee g) \leftrightarrow (\neg f \wedge \neg g)$$

zur Anwendung. Wir beschreiben die Umformung in Negations-Normalform durch die folgenden Gleichungen:

1. $\mathbf{nnf}(\neg f) = \mathbf{neg}(f)$,
2. $\mathbf{nnf}(f_1 \wedge f_2) = \mathbf{nnf}(f_1) \wedge \mathbf{nnf}(f_2)$,
3. $\mathbf{nnf}(f_1 \vee f_2) = \mathbf{nnf}(f_1) \vee \mathbf{nnf}(f_2)$.

Die Hilfsprozedur **neg**, die die Negations-Normalform von $\neg f$ berechnet, spezifizieren wir ebenfalls durch rekursive Gleichungen:

1. $\mathbf{neg}(p) = \mathbf{nnf}(\neg p) = \neg p$ für alle Aussage-Variablen p ,
2. $\mathbf{neg}(\neg f) = \mathbf{nnf}(\neg \neg f) = \mathbf{nnf}(f)$,
3.
$$\begin{aligned} \mathbf{neg}(f_1 \wedge f_2) &= \mathbf{nnf}(\neg(f_1 \wedge f_2)) \\ &= \mathbf{nnf}(\neg f_1 \vee \neg f_2) \\ &= \mathbf{nnf}(\neg f_1) \vee \mathbf{nnf}(\neg f_2) \\ &= \mathbf{neg}(f_1) \vee \mathbf{neg}(f_2), \end{aligned}$$
4.
$$\begin{aligned} \mathbf{neg}(f_1 \vee f_2) &= \mathbf{nnf}(\neg(f_1 \vee f_2)) \\ &= \mathbf{nnf}(\neg f_1 \wedge \neg f_2) \\ &= \mathbf{nnf}(\neg f_1) \wedge \mathbf{nnf}(\neg f_2) \\ &= \mathbf{neg}(f_1) \wedge \mathbf{neg}(f_2). \end{aligned}$$

Als letztes stellen wir die Prozeduren vor, mit denen die Formeln, die bereits in Negations-Normalform sind, ausmultipliziert und dadurch in konjunktive Normalform gebracht werden. Gleichzeitig werden die zu normalisierende Formel dabei in die Mengen-Schreibweise transformiert, d.h. die Formeln werden als Mengen von Mengen von Literalen dargestellt. Dabei interpretieren wir eine Menge von Literalen als Disjunktion der Literale und eine Menge von Klauseln interpretieren wir als Konjunktion der Klauseln. Abbildung 5.7 auf Seite 99 zeigt die Implementierung der Funktion **knf**.

1. Falls die Formel f , die wir in KNF transformieren wollen, die Form

$$f = \neg g$$

hat, so muss g eine Aussage-Variable sein, denn f ist ja bereits in Negations-Normalform. Damit können wir f in eine Klausel transformieren, indem wir $\{\neg g\}$, also $\{f\}$ schreiben. Da eine KNF eine Menge von Klauseln ist, ist dann $\{\{f\}\}$ das Ergebnis, das wir in Zeile 3 zurück geben.

2. Falls $f = f_1 \wedge f_2$ ist, transformieren wir zunächst f_1 und f_2 in KNF. Dabei erhalten wir

$$\mathbf{knf}(f_1) = \{h_1, \dots, h_m\} \quad \text{und} \quad \mathbf{knf}(f_2) = \{k_1, \dots, k_n\}.$$

Dabei sind die h_i und die k_j Klauseln. Um nun die KNF von $f_1 \wedge f_2$ zu bilden, reicht es aus, die Vereinigung dieser beiden Mengen zu bilden, wir haben also

$$\mathbf{knf}(f_1 \wedge f_2) = \mathbf{knf}(f_1) \cup \mathbf{knf}(f_2).$$

Das liefert Zeile 4 der Implementierung.

```

1  nnf := procedure(f) {
2      match (f) {
3          case !g      : return neg(g);
4          case g && h : return nnf(g) && nnf(h);
5          case g || h : return nnf(g) || nnf(h);
6          default      : return f;
7      }
8  };
9  neg := procedure(f) {
10     match (f) {
11         case !g      : return nnf(g);
12         case g && h : return neg(g) || neg(h);
13         case g || h : return neg(g) && neg(h);
14         default      : return !f;
15     }
16 };

```

Abbildung 5.6: Berechnung der Negations-Normalform.

3. Falls $f = f_1 \vee f_2$ ist, transformieren wir zunächst f_1 und f_2 in KNF. Dabei erhalten wir

$$\mathbf{knf}(f_1) = \{h_1, \dots, h_m\} \quad \text{und} \quad \mathbf{knf}(f_2) = \{k_1, \dots, k_n\}.$$

Dabei sind die h_i und die k_j Klauseln. Um nun die KNF von $f_1 \vee f_2$ zu bilden, rechnen wir wie folgt:

$$\begin{aligned} & f_1 \vee f_2 \\ \Leftrightarrow & (h_1 \wedge \cdots \wedge h_m) \vee (k_1 \wedge \cdots \wedge k_n) \\ \Leftrightarrow & (h_1 \vee k_1) \quad \wedge \quad \cdots \quad \wedge \quad (h_m \vee k_1) \quad \wedge \\ & \qquad \vdots \qquad \qquad \qquad \qquad \qquad \qquad \vdots \\ & (h_1 \vee k_n) \quad \wedge \quad \cdots \quad \wedge \quad (h_m \vee k_n) \\ \Leftrightarrow & \{h_i \vee k_j : i \in \{1, \dots, m\}, j \in \{1, \dots, n\}\} \end{aligned}$$

Berücksichtigen wir noch, dass Klauseln in der Mengen-Schreibweise als Mengen von Literalen aufgefaßt werden, die implizit disjunktiv verknüpft werden, so können wir für $h_i \vee k_j$ auch $h_i \cup k_j$ schreiben. Insgesamt erhalten wir damit

$$\mathbf{k}\mathbf{n}\mathbf{f}(f_1 \vee f_2) = \{h \cup k \mid h \in \mathbf{k}\mathbf{n}\mathbf{f}(f_1) \wedge k \in \mathbf{k}\mathbf{n}\mathbf{f}(f_2)\}.$$

Das liefert die Zeile 5 der Implementierung der Prozedur `knf`.

4. Falls die Formel f , die wir in KNF transformieren wollen, eine Aussage-Variable ist, so transformieren wir f zunächst in eine Klausel. Das liefert $\{f\}$. Da eine KNF eine Menge von Klauseln ist, ist die KNF dann $\{\{f\}\}$. Dieses Ergebnis geben wir in Zeile 6 zurück.

Zum Abschluß zeigen wir in Abbildung 5.8 auf Seite 99 wie die einzelnen Funktionen zusammenspielen. Die Funktion `normalize` eliminiert zunächst die Junktoren “ \leftrightarrow ” und “ \rightarrow ” und bringt die Formel in Negations-Normalform. Die Negations-Normalform wird nun mit Hilfe der Funktion `knf` in konjunktive Normalform gebracht, wobei gleichzeitig die Formel in Mengen-Schreibweise überführt wird. Im letzten Schritt entfernt die Funktion `simplify` alle Klauseln, die trivial sind. Eine Klausel k ist dann trivial, wenn es eine aussagenlogische Variablen p gibt, so dass sowohl p als auch $\neg p$ in k der Menge k auftritt. Daher berechnet die Funktion `isTrivial` für eine Klausel c zunächst die Menge

```
{ p in c | fct(p) == "^variable" }
```

```

1  knf := procedure(f) {
2      match (f) {
3          case !g      : return { { !g } };
4          case g && h : return knf(g) + knf(h);
5          case g || h : return { k1 + k2 : k1 in knf(g), k2 in knf(h) };
6          default      : return { { f } }; // f is a variable
7      }
8  };

```

Abbildung 5.7: Berechnung der konjunktiven Normalform.

aller Variablen, die in der Klausel c auftreten. Anschließend wird die Menge

$$\{ \text{args}(l)[1] : l \text{ in } c \mid \text{fct}(l) == \text{"^negation"} \}$$

berechnet. Diese Menge enthält alle die aussagenlogischen Variablen p , für die $\neg p$ ein Element der Klausel c ist, denn die Formel $\neg p$ wird intern in SETLX durch den Term

$$\text{^negation}(p)$$

dargestellt. Falls also das Literal l die Form $l = \neg p$ hat, so hat der Term l das Funktions-Zeichen ^negation . Die Anwendung der Funktion fct auf l liefert uns genau dieses Funktions-Zeichen. Der Ausdruck $\text{args}(l)$ berechnet die Liste der Argumente des Funktions-Zeichens ^negation und diese Liste enthält als einziges Element gerade die aussagenlogische Variable p , die wir daher durch den Ausdruck

$$\text{args}(l)[1]$$

aus dem Literal l extrahieren können. Falls nun die beiden oben berechneten Mengen ein gemeinsames Element haben, so ist die Klausel c trivial. Dies wird dadurch geprüft, dass der Schnitt der beiden Mengen berechnet wird. Das vollständige Programm zur Berechnung der konjunktiven Normalform finden Sie als die Datei `knf.stlx` auf meiner Webseite.

```

1  normalize := procedure(f) {
2      n1 := elimGdw(f);
3      n2 := elimFolgt(n1);
4      n3 := nnf(n2);
5      n4 := knf(n3);
6      return simplify(n4);
7  };
8  simplify := procedure(k) {
9      return { c : c in k | !isTrivial(c) };
10 };
11 isTrivial := procedure(c) {
12     return { p : p in c | fct(p) == "^variable" } *
13         { args(l)[1] : l in c | fct(l) == "^not" } != {};
14 };

```

Abbildung 5.8: Normalisierung einer Formel

5.5 Der Herleitungs-Begriff

Ist $\{f_1, \dots, f_n\}$ eine Menge von Formeln, und g eine weitere Formel, so können wir uns fragen, ob die Formel g aus f_1, \dots, f_n folgt, ob also

$$\models f_1 \wedge \dots \wedge f_n \rightarrow g$$

gilt. Es gibt verschiedene Möglichkeiten, diese Frage zu beantworten. Ein Verfahren kennen wir schon: Zunächst überführen wir die Formel $f_1 \wedge \dots \wedge f_n \rightarrow g$ in konjunktive Normalform. Wir erhalten dann eine Menge $\{k_1, \dots, k_n\}$ von Klauseln, deren Konjunktion zu der Formel

$$f_1 \wedge \dots \wedge f_n \rightarrow g$$

äquivalent ist. Diese Formel ist nun genau dann eine Tautologie, wenn jede der Klauseln k_1, \dots, k_n trivial ist.

Das oben dargestellte Verfahren ist aber sehr aufwendig. Wir zeigen dies an Hand eines Beispiels und wenden das Verfahren an, um zu entscheiden, ob $p \rightarrow r$ aus den beiden Formeln $p \rightarrow q$ und $q \rightarrow r$ folgt. Wir bilden also die konjunktive Normalform der Formel

$$h := (p \rightarrow q) \wedge (q \rightarrow r) \rightarrow p \rightarrow r$$

und erhalten

$$(p \vee \neg p \vee r \vee \neg r) \wedge (\neg q \vee \neg p \vee r \vee \neg r) \wedge (\neg q \vee \neg p \vee q \vee r) \wedge (p \vee \neg p \vee q \vee r).$$

Zwar können wir jetzt sehen, dass die Formel h eine Tautologie ist, aber angesichts der Tatsache, dass wir mit bloßem Auge sehen, dass $p \rightarrow r$ aus den Formeln $p \rightarrow q$ und $q \rightarrow r$ folgt, ist die Rechnung doch sehr mühsam.

Wir stellen daher nun ein weiteres Verfahren vor, mit dessen Hilfe wir entscheiden können, ob eine Formel aus einer gegebenen Menge von Formeln folgt. Die Idee bei diesem Verfahren ist es, die Formel f mit Hilfe von *Schluss-Regeln* aus den gegebenen Formeln f_1, \dots, f_n herzuleiten. Das Konzept einer Schluss-Regel wird in der nun folgenden Definition festgelegt.

Definition 30 (Schluss-Regel) Eine *Schluss-Regel* ist eine Paar $\langle \{f_1, \dots, f_n\}, k \rangle$. Dabei ist $\{f_1, \dots, f_n\}$ eine Menge von Formeln und k ist eine einzelne Formel. Die Formeln f_1, \dots, f_n bezeichnen wir als *Prämissen*, die Formel k heißt die *Konklusion* der Schluss-Regel. Ist das Paar $\langle \{f_1, \dots, f_n\}, k \rangle$ eine Schluss-Regel, so schreiben wir dies als:

$$\frac{f_1 \quad \dots \quad f_n}{k}.$$

Wir lesen diese Schluss-Regel wie folgt: “Aus f_1, \dots, f_n kann auf k geschlossen werden.” □

Beispiele für Schluss-Regeln:

<i>Modus Ponens</i>	<i>Modus Ponendo Tollens</i>	<i>Modus Tollendo Tollens</i>
$\frac{p \quad p \rightarrow q}{q}$	$\frac{\neg q \quad p \rightarrow q}{\neg p}$	$\frac{\neg p \quad p \rightarrow q}{\neg q}$

Die Definition der Schluss-Regel schränkt zunächst die Formeln, die als Prämissen bzw. Konklusion verwendet werden können, nicht weiter ein. Es ist aber sicher nicht sinnvoll, beliebige Schluss-Regeln zuzulassen. Wollen wir Schluss-Regeln in Beweisen verwenden, so sollten die Schluss-Regeln in dem in der folgenden Definition erklärten Sinne *korrekt* sein.

Definition 31 (Korrekte Schluss-Regel) Eine Schluss-Regel

$$\frac{f_1 \quad \dots \quad f_n}{k}$$

ist genau dann *korrekt*, wenn $\models f_1 \wedge \dots \wedge f_n \rightarrow k$ gilt. \square

Mit dieser Definition sehen wir, dass die oben als “*Modus Ponens*” und “*Modus Ponendo Tollens*” bezeichneten Schluss-Regeln korrekt sind, während die als “*Modus Tollendo Tollens*” bezeichnete Schluss-Regel nicht korrekt ist.

Im Folgenden gehen wir davon aus, dass alle Formeln Klauseln sind. Einerseits ist dies keine echte Einschränkung, denn wir können ja jede Formel in eine äquivalente Menge von Klauseln umrechnen. Andererseits haben viele in der Praxis auftretende aussagenlogische Probleme die Gestalt von Klauseln. Daher stellen wir jetzt eine Schluss-Regel vor, in der sowohl die Prämissen als auch die Konklusion Klauseln sind.

Definition 32 (Schnitt-Regel) Ist p eine aussagenlogische Variable und sind k_1 und k_2 Mengen von Literalen, die wir als Klauseln interpretieren, so bezeichnen wir die folgende Schluss-Regel als die *Schnitt-Regel*:

$$\frac{k_1 \cup \{p\} \quad \{\neg p\} \cup k_2}{k_1 \cup k_2}.$$

\diamond

Die Schnitt-Regel ist sehr allgemein. Setzen wir in der obigen Definition für $k_1 = \{\}$ und $k_2 = \{q\}$ ein, so erhalten wir die folgende Regel als Spezialfall:

$$\frac{\{\} \cup \{p\} \quad \{\neg p\} \cup \{q\}}{\{\} \cup \{q\}}$$

Interpretieren wir nun die Mengen als Disjunktionen, so haben wir:

$$\frac{p \quad \neg p \vee q}{q}$$

Wenn wir jetzt noch berücksichtigen, dass die Formel $\neg p \vee q$ äquivalent ist zu der Formel $p \rightarrow q$, dann ist das nichts anderes als der *Modus Ponens*. Die Regel *Modus Tollens* ist ebenfalls ein Spezialfall der Schnitt-Regel. Wir erhalten diese Regel, wenn wir in der Schnitt-Regel $k_1 = \{\neg q\}$ und $k_2 = \{\}$ setzen.

Satz 33 Die Schnitt-Regel ist korrekt.

Beweis: Wir müssen zeigen, dass

$$\models (k_1 \vee p) \wedge (\neg p \vee k_2) \rightarrow k_1 \vee k_2$$

gilt. Dazu überführen wir die obige Formel in konjunktive Normalform:

$$\begin{aligned} & (k_1 \vee p) \wedge (\neg p \vee k_2) \rightarrow k_1 \vee k_2 \\ \Leftrightarrow & \neg((k_1 \vee p) \wedge (\neg p \vee k_2)) \vee k_1 \vee k_2 \\ \Leftrightarrow & \neg(k_1 \vee p) \vee \neg(\neg p \vee k_2) \vee k_1 \vee k_2 \\ \Leftrightarrow & (\neg k_1 \wedge \neg p) \vee (p \wedge \neg k_2) \vee k_1 \vee k_2 \\ \Leftrightarrow & (\neg k_1 \vee p \vee k_1 \vee k_2) \wedge (\neg k_1 \vee \neg k_2 \vee k_1 \vee k_2) \wedge (\neg p \vee p \vee k_1 \vee k_2) \wedge (\neg p \vee \neg k_2 \vee k_1 \vee k_2) \\ \Leftrightarrow & \top \wedge \top \wedge \top \wedge \top \\ \Leftrightarrow & \top \end{aligned}$$

\square

Definition 34 (\vdash) Es sei M eine Menge von Klauseln und f sei eine einzelne Klausel. Die Formeln aus M bezeichnen wir als unsere Annahmen. Unser Ziel ist es, mit diesen Annahmen die Formel f zu beweisen. Dazu definieren wir induktiv die Relation

$$M \vdash f.$$

Wir lesen “ $M \vdash f$ ” als “ M leitet f her”. Die induktive Definition ist wie folgt:

1. Aus einer Menge M von Annahmen kann jede der Annahmen hergeleitet werden:

Falls $f \in M$ ist, dann gilt $M \vdash f$.

2. Sind $k_1 \cup \{p\}$ und $\{\neg p\} \cup k_2$ Klauseln, die aus M hergeleitet werden können, so kann mit der Schnitt-Regel auch die Klausel $k_1 \cup k_2$ aus M hergeleitet werden:

Falls sowohl $M \vdash k_1 \cup \{p\}$ als auch $M \vdash \{\neg p\} \cup k_2$ gilt, dann gilt auch $M \vdash k_1 \cup k_2$.

Dies ist die Schnitt-Regel. □

Beispiel: Um den Beweis-Begriff zu veranschaulichen geben wir ein Beispiel und zeigen

$$\{ \{\neg p, q\}, \{\neg q, \neg p\}, \{\neg q, p\}, \{q, p\} \} \vdash \perp.$$

Gleichzeitig zeigen wir an Hand des Beispiels, wie wir Beweise zu Papier bringen:

1. Aus $\{\neg p, q\}$ und $\{\neg q, \neg p\}$ folgt mit der Schnitt-Regel $\{\neg p, \neg p\}$. Wegen $\{\neg p, \neg p\} = \{\neg p\}$ schreiben wir dies als

$$\{\neg p, q\}, \{\neg q, \neg p\} \vdash \{\neg p\}.$$

Dieses Beispiel zeigt, dass die Klausel $k_1 \cup k_2$ durchaus auch weniger Elemente enthalten kann als die Summe $\#k_1 + \#k_2$. Dieser Fall tritt genau dann ein, wenn es Literale gibt, die sowohl in k_1 als auch in k_2 vorkommen.

2. $\{\neg q, \neg p\}, \{p, \neg q\} \vdash \{\neg q\}$.
 3. $\{p, q\}, \{\neg q\} \vdash \{p\}$.
 4. $\{\neg p\}, \{p\} \vdash \{\}$.

Als weiteres Beispiel zeigen wir nun, dass $p \rightarrow r$ aus $p \rightarrow q$ und $q \rightarrow r$ folgt. Dazu überführen wir zunächst alle Formeln in Klauseln:

$$\text{knf}(p \rightarrow q) = \{ \{\neg p, q\} \}, \quad \text{knf}(q \rightarrow r) = \{ \{\neg q, r\} \}, \quad \text{knf}(p \rightarrow r) = \{ \{\neg p, r\} \}.$$

Wir haben also $M = \{ \{\neg p, q\}, \{\neg q, r\} \}$ und müssen zeigen, dass

$$M \vdash \{\neg p, r\}$$

folgt. Der Beweis besteht aus einer einzigen Zeile:

$$\{\neg p, q\}, \{\neg q, r\} \vdash \{\neg p, r\}.$$

5.5.1 Eigenschaften des Herleitungs-Begriffs

Die Relation \vdash hat zwei wichtige Eigenschaften:

Satz 35 (Korrektheit) Ist $\{k_1, \dots, k_n\}$ eine Menge von Klauseln und k eine einzelne Klausel, so haben wir:

$$\text{Wenn } \{k_1, \dots, k_n\} \vdash k \text{ gilt, dann gilt auch } \models k_1 \wedge \dots \wedge k_n \rightarrow k.$$

Beweis: Der Beweis verläuft durch eine Induktion nach der Definition der Relation \vdash .

1. Fall: Es gilt $\{k_1, \dots, k_n\} \vdash k$ weil $k \in \{k_1, \dots, k_n\}$ ist. Dann gibt es also ein $i \in \{1, \dots, n\}$, so dass $k = k_i$ ist. In diesem Fall müssen wir

$$\models k_1 \wedge \dots \wedge k_n \rightarrow k_i$$

zeigen, was offensichtlich ist.

2. Fall: Es gilt $\{k_1, \dots, k_n\} \vdash k$ weil es eine aussagenlogische Variable p und Klauseln g und h gibt, so dass

$$\{k_1, \dots, k_n\} \vdash g \cup \{p\} \quad \text{und} \quad \{k_1, \dots, k_n\} \vdash h \cup \{\neg p\}$$

gilt und daraus haben wir mit der Schnitt-Regel auf

$$\{k_1, \dots, k_n\} \vdash g \cup h$$

geschlossen mit $k = g \cup h$.

Nach Induktions-Voraussetzung haben wir dann

$$\models k_1 \wedge \dots \wedge k_n \rightarrow g \vee p \quad \text{und} \quad \models k_1 \wedge \dots \wedge k_n \rightarrow h \vee \neg p.$$

Wegen

$$\models (g \vee p) \wedge (h \vee \neg p) \rightarrow g \vee h \quad \text{und} \quad k = g \cup h$$

folgt daraus die Behauptung. □

Die Umkehrung dieses Satzes gilt leider nur in abgeschwächter Form und zwar dann, wenn k die leere Klausel ist, also im Fall $k = \perp$.

Satz 36 (Widerlegungs-Vollständigkeit) Ist $\{k_1, \dots, k_n\}$ eine Menge von Klauseln, so haben wir:

Wenn $\models k_1 \wedge \dots \wedge k_n \rightarrow \perp$ gilt, dann gilt auch $\{k_1, \dots, k_n\} \vdash \{\}$.

Diesen Satz im Rahmen der Vorlesung zu beweisen würde zuviel Zeit kosten. Ein Beweis findet sich beispielsweise in dem Buch von Schöning [\[Sch87\]](#).

5.6 Das Verfahren von Davis und Putnam

In der Praxis stellt sich oft die Aufgabe, für eine gegebene Menge von Klauseln K eine Belegung \mathcal{I} der Variablen zu berechnen, so dass

$$\text{eval}(k, \mathcal{I}) = \text{true} \quad \text{für alle } k \in K$$

gilt. In diesem Fall sagen wir auch, dass die Belegung \mathcal{I} eine *Lösung* der Klausel-Menge K ist. Wir werden in diesem Abschnitt ein Verfahren vorstellen, mit dem eine solche Aufgabe bearbeitet werden kann. Dieses Verfahren geht auf Davis und Putnam [\[DLL62\]](#) zurück. Verfeinerungen dieses Verfahrens werden beispielsweise eingesetzt, um die Korrektheit digitaler elektronischer Schaltungen nachzuweisen.

Um das Verfahren zu motivieren überlegen wir zunächst, bei welcher Form der Klausel-Menge K unmittelbar klar ist, ob es eine Belegung gibt, die K löst und wie diese Belegung aussieht. Betrachten wir dazu ein Beispiel:

$$K_1 = \{ \{p\}, \{\neg q\}, \{r\}, \{\neg s\}, \{\neg t\} \}$$

Die Klausel-Menge K_1 entspricht der aussagenlogischen Formel

$$p \wedge \neg q \wedge r \wedge \neg s \wedge \neg t.$$

Daher ist K_1 lösbar und die Belegung

$$\mathcal{I} = \{ \langle p, \text{true} \rangle, \langle q, \text{false} \rangle, \langle r, \text{true} \rangle, \langle s, \text{false} \rangle, \langle t, \text{false} \rangle \}$$

ist eine Lösung. Betrachten wir ein weiteres Beispiel:

$$K_2 = \{ \{\}, \{p\}, \{\neg q\}, \{r\} \}$$

Diese Klausel-Menge entspricht der Formel

$$\perp \wedge p \wedge \neg q \wedge r.$$

Offensichtlich ist K_2 unlösbar. Als letztes Beispiel betrachten wir

$$K_3 = \{\{p\}, \{\neg q\}, \{\neg p\}\}.$$

Diese Klausel-Menge kodiert die Formel

$$p \wedge \neg q \wedge \neg p$$

Offenbar ist K_3 ebenfalls unlösbar, denn eine Lösung \mathcal{I} müßte p gleichzeitig wahr und falsch machen. Wir nehmen die an den letzten drei Beispielen gemachten Beobachtungen zum Anlaß für zwei Definitionen.

Definition 37 (Unit-Klausel) Eine Klausel k heißt *Unit-Klausel*, wenn k nur aus einem Literal besteht. Es gilt dann

$$k = \{p\} \quad \text{oder} \quad k = \{\neg p\}$$

für eine Aussage-Variable p .

Definition 38 (Triviale Klausel-Mengen) Eine Klausel-Menge K heißt *trivial* wenn einer der beiden folgenden Fälle vorliegt.

1. K enthält die leere Klausel: $\{\} \in K$.

In diesem Fall ist K offensichtlich unlösbar.

2. K enthält nur Unit-Klausel mit verschiedenen Aussage-Variablen. Bezeichnen wir die Menge der aussagenlogischen Variablen mit \mathcal{P} , so schreibt sich diese Bedingung als

$$\forall k \in K : \text{card}(k) = 1 \quad \text{und} \quad \forall p \in \mathcal{P} : \neg(\{p\} \in K \wedge \{\neg p\} \in K).$$

Dann ist

$$\mathcal{I} = \{\langle p, \text{true} \rangle \mid \{p\} \in K\} \cup \{\langle p, \text{false} \rangle \mid \{\neg p\} \in K\}$$

eine Lösung von K .

Wie können wir nun eine Menge von Klauseln so vereinfachen, dass die Menge schließlich nur noch aus Unit-Klauseln besteht? Es gibt drei Möglichkeiten, Klauselmengen zu vereinfachen:

1. Schnitt-Regel,
2. Subsumption und
3. Fallunterscheidung.

Wir betrachten diese Möglichkeiten jetzt der Reihe nach.

5.6.1 Vereinfachung mit der Schnitt-Regel

Eine typische Anwendung der Schnitt-Regel hat die Form:

$$\frac{k_1 \cup \{p\} \quad \{\neg p\} \cup k_2}{k_1 \cup k_2}$$

Die hierbei erzeugte Klausel $k_1 \cup k_2$ wird in der Regel mehr Literale enthalten als die Prämissen $k_1 \cup \{p\}$ und $\{\neg p\} \cup k_2$. Enthält die Klausel $k_1 \cup \{p\}$ insgesamt $m+1$ Literale und enthält die Klausel $\{\neg p\} \cup k_2$ insgesamt $n+1$ Literale, so kann die Konklusion $k_1 \cup k_2$ insgesamt $m+n$ Literale enthalten. Natürlich können es auch weniger Literale sein, und zwar dann, wenn es Literale gibt, die sowohl in k_1 als auch in k_2 auftreten. Im allgemeinen ist $m+n$ größer als $m+1$ und als $n+1$. Die Klauseln wachsen nur dann sicher nicht, wenn entweder $n=0$ oder $m=0$ ist. Dieser Fall liegt vor, wenn einer der beiden Klauseln nur aus einem Literal besteht und folglich eine *Unit-Klausel* ist. Da es unser Ziel ist, die Klausel-Mengen zu vereinfachen, lassen wir nur solche Anwendungen

der Schnitt-Regel zu, bei denen eine der Klausel eine Unit-Klausel ist. Solche Schnitte bezeichnen wir als *Unit-Schnitte*. Um alle mit einer gegebenen Unit-Klausel $\{l\}$ möglichen Schnitte durchführen zu können, definieren wir eine Funktion

$$\text{unitCut} : 2^{\mathcal{K}} \times \mathcal{L} \rightarrow 2^{\mathcal{K}}$$

so, dass für eine Klausel-Menge K und eine Unit-Klausel $\{l\}$ die Funktion $\text{unitCut}(K, l)$ die Klausel-Menge K soweit wie möglich mit Unit-Schnitten mit der Klausel $\{l\}$ vereinfacht:

$$\text{unitCut}(K, l) = \left\{ k \setminus \{\overline{l}\} \mid k \in K \right\}$$

5.6.2 Vereinfachung durch Subsumption

Das Prinzip der Subsumption demonstrieren wir zunächst an einem Beispiel. Wir betrachten

$$K = \{\{p, q, \neg r\}, \{p\}\} \cup M.$$

Offenbar impliziert die Klausel $\{p\}$ die Klausel $\{p, q, \neg r\}$, immer wenn $\{p\}$ erfüllt ist, ist automatisch auch $\{q, p, \neg r\}$ erfüllt, denn es gilt

$$\models p \rightarrow q \vee p \vee \neg r.$$

Allgemein sagen wir, dass eine Klausel k von einer Unit-Klausel u *subsumiert* wird, wenn

$$u \subseteq k$$

gilt. Ist K eine Klausel-Menge mit $k \in K$ und $u \in K$ und wird k durch u subsumiert, so können wir K durch Unit-Subsumption zu $K - \{k\}$ vereinfachen, indem wir die Klausel k aus K löschen. Allgemein definieren wir eine Funktion

$$\text{subsume} : 2^{\mathcal{K}} \times \mathcal{L} \rightarrow 2^{\mathcal{K}}$$

die eine gegebene Klauselmeng K , die die Unit-Klausel $\{l\}$ enthält, mittels Subsumption dadurch vereinfacht, dass alle durch $\{l\}$ subsumierten Klauseln aus K gelöscht werden. Die Unit-Klausel $\{l\}$ behalten wir natürlich. Daher definieren wir:

$$\text{subsume}(K, l) := (K \setminus \{k \in K \mid l \in k\}) \cup \{\{l\}\} = \{k \in K \mid l \notin k\} \cup \{\{l\}\}.$$

In der obigen Definition muss $\{l\}$ in das Ergebnis eingefügt werden, weil die Menge $\{k \in K \mid l \notin k\}$ die Unit-Klausel $\{l\}$ nicht enthält.

5.6.3 Vereinfachung durch Fallunterscheidung

Ein Kalkül, der nur mit Unit-Schnitten und Subsumption arbeitet, ist nicht widerlegungs-vollständig. Wir brauchen daher eine weitere Möglichkeit, Klausel-Mengen zu vereinfachen. Eine solche Möglichkeit bietet das Prinzip der *Fallunterscheidung*. Dieses Prinzip basiert auf dem folgenden Satz.

Satz 39 Ist K eine Menge von Klauseln und ist p eine aussagenlogische Variable, so ist K genau dann erfüllbar, wenn $K \cup \{\{p\}\}$ oder $K \cup \{\{\neg p\}\}$ erfüllbar ist.

Beweis: Ist K erfüllbar durch eine Belegung \mathcal{I} , so gibt es für $\mathcal{I}(p)$ zwei Möglichkeiten: Falls $\mathcal{I}(p) = \text{true}$ ist, ist damit auch die Menge $K \cup \{\{p\}\}$ erfüllbar, andernfalls ist $K \cup \{\{\neg p\}\}$ erfüllbar.

Da K sowohl eine Teilmenge von $K \cup \{\{p\}\}$ als auch von $K \cup \{\{\neg p\}\}$ ist, ist klar, dass K erfüllbar ist, wenn eine dieser Mengen erfüllbar sind. \square

Wir können nun eine Menge K von Klauseln dadurch vereinfachen, dass wir eine aussagenlogische Variable p wählen, die in K vorkommt. Anschließend bilden wir die Mengen

$$K_1 := K \cup \{\{p\}\} \quad \text{und} \quad K_2 := K \cup \{\{\neg p\}\}$$

und untersuchen rekursiv ob K_1 erfüllbar ist. Falls wir eine Lösung für K_1 finden, ist dies auch eine Lösung für die ursprüngliche Klausel-Menge K und wir haben unser Ziel erreicht. Andernfalls untersuchen wir rekursiv ob

K_2 erfüllbar ist. Falls wir nun eine Lösung finden, ist dies auch eine Lösung von K und wenn wir für K_2 keine Lösung finden, dann hat auch K keine Lösung. Die rekursive Untersuchung von K_1 bzw. K_2 ist leichter, weil ja wir dort mit den Unit-Klausel $\{p\}$ bzw. $\{\neg p\}$ zunächst Unit-Subsumption und anschließend Unit-Schnitte durchführen können.

5.6.4 Der Algorithmus

Wir können jetzt den Algorithmus von Davis und Putnam skizzieren. Gegeben sei eine Menge K von Klauseln. Gesucht ist dann eine Lösung von K . Wir suchen also eine Belegung \mathcal{I} , so dass gilt:

$$\mathcal{I}(k) = \text{true} \quad \text{für alle } k \in K.$$

Das Verfahren von Davis und Putnam besteht nun aus den folgenden Schritten.

1. Führe alle Unit-Schnitte aus, die mit Klauseln aus K möglich sind und führe zusätzlich alle Unit-Subsumptionen aus.
2. Falls K nun trivial ist, sind wir fertig.
3. Andernfalls wählen wir eine aussagenlogische Variable p , die in K auftritt.

- (a) Jetzt versuchen wir rekursiv die Klausel-Menge

$$K \cup \{\{p\}\}$$

zu lösen. Falls diese gelingt, haben wir eine Lösung von K .

- (b) Andernfalls versuchen wir die Klausel-Menge

$$K \cup \{\{\neg p\}\}$$

zu lösen. Wenn auch dies fehlschlägt, ist K unlösbar, andernfalls haben wir eine Lösung von K .

Für die Implementierung ist es zweckmäßig, die beiden oben definierten Funktionen *unitCut()* und *subsume()* zusammen zu fassen. Wir definieren eine Funktion

$$\text{reduce} : 2^{\mathcal{K}} \times \mathcal{L} \rightarrow 2^{\mathcal{K}}$$

wie folgt:

$$\text{reduce}(K, l) = \{k \setminus \{\bar{l}\} \mid k \in K \wedge \bar{l} \in k\} \cup \{k \in K \mid \bar{l} \notin k \wedge l \notin k\} \cup \{\{l\}\}.$$

Die Menge enthält also einerseits die Ergebnisse von Schnitten mit der Unit-Klausel $\{l\}$ und andererseits nur noch die Klauseln k , die mit l nichts zu tun haben weil weder $l \in k$ noch $\bar{l} \in k$ gilt. Außerdem fügen wir auch noch die Unit-Klausel $\{l\}$ hinzu. Dadurch erreichen wir, dass die beiden Mengen K und $\text{reduce}(K, l)$ logisch äquivalent sind, wenn wir dieses Mengen als Formeln in konjunktiver Normalform interpretieren.

5.6.5 Ein Beispiel

Zur Veranschaulichung demonstrieren wir das Verfahren von Davis und Putnam an einem Beispiel. Die Menge K sei wie folgt definiert:

$$K := \left\{ \{p, q, s\}, \{\neg p, r, \neg t\}, \{r, s\}, \{\neg r, q, \neg p\}, \right. \\ \left. \{\neg s, p\}, \{\neg p, \neg q, s, \neg r\}, \{p, \neg q, s\}, \{\neg r, \neg s\}, \{\neg p, \neg s\} \right\}.$$

Wir zeigen nun mit dem Verfahren von Davis und Putnam, dass K nicht lösbar ist. Da die Menge K keine Unit-Klauseln enthält, ist im ersten Schritt nichts zu tun. Da K nicht trivial ist, sind wir noch nicht fertig. Also gehen wir jetzt zu Schritt 3 und wählen eine aussagenlogische Variable, die in K auftritt. An dieser Stelle ist es sinnvoll eine Variable zu wählen, die in möglichst vielen Klauseln von K auftritt. Wir wählen daher die aussagenlogische Variable p .

1. Zunächst bilden wir jetzt die Menge

$$K_0 := K \cup \{\{p\}\}$$

und versuchen, diese Menge zu lösen. Dazu bilden wir jetzt

$$K_1 := \text{reduce}(K_0, p) = \left\{ \{r, \neg t\}, \{r, s\}, \{\neg r, q\}, \{\neg q, s, \neg r\}, \{\neg r, \neg s\}, \{\neg s\}, \{p\} \right\}.$$

Die Klausel-Menge K_1 enthält die Unit-Klausel $\{\neg s\}$, so dass wir als nächstes mit dieser Klausel reduzieren können:

$$K_2 := \text{reduce}(K_1, \neg s) = \left\{ \{r, \neg t\}, \{r\}, \{\neg r, q\}, \{\neg q, \neg r\}, \{\neg s\}, \{p\} \right\}.$$

Hier haben wir nun die neue Unit-Klausel $\{r\}$, mit der wir als nächstes reduzieren:

$$K_3 := \text{reduce}(K_2, r) = \left\{ \{r\}, \{q\}, \{\neg q\}, \{\neg s\}, \{p\} \right\}$$

Da K_3 die Unit-Klausel $\{q\}$ enthält, reduzieren wir jetzt mit q :

$$K_4 := \text{reduce}(K_3, q) = \left\{ \{r\}, \{q\}, \{\}, \{\neg s\}, \{p\} \right\}.$$

Die Klausel-Menge K_4 enthält die leere Klausel und ist damit unlösbar.

2. Also bilden wir jetzt die Menge

$$K_5 := K \cup \{\{\neg p\}\}$$

und versuchen, diese Menge zu lösen. Dazu bilden wir

$$K_6 = \text{reduce}(K_5, \neg p) = \left\{ \{q, s\}, \{r, s\}, \{\neg s\}, \{\neg q, s\}, \{\neg r, \neg s\}, \{\neg p\} \right\}.$$

Die Menge K_6 enthält die Unit-Klausel $\{\neg s\}$. Wir bilden daher

$$K_7 = \text{reduce}(K_6, \neg s) = \left\{ \{q\}, \{r\}, \{\neg s\}, \{\neg q\}, \{\neg p\} \right\}.$$

Die Menge K_7 enthält die neue Unit-Klausel $\{q\}$, mit der wir als nächstes reduzieren:

$$K_8 = \text{reduce}(K_7, q) = \left\{ \{q\}, \{r\}, \{\neg s\}, \{\}, \{\neg p\} \right\}.$$

Da K_8 die leere Klausel enthält, ist K_8 und damit auch die ursprünglich gegebene Menge K unlösbar.

5.6.6 Implementierung des Algorithmus von Davis und Putnam

Wir zeigen jetzt die Implementierung der Prozedur `davisPutnam`, mit der die Frage, ob eine Menge von Klauseln erfüllbar ist, beantwortet werden kann. Die Implementierung ist in Abbildung 5.9 auf Seite 108 gezeigt. Die Prozedur erhält zwei Argumente: `clauses` und `literals`. `clauses` ist eine Menge von Klauseln und `literals` ist eine Menge von Literalen. Falls die Vereinigung dieser beiden Mengen erfüllbar ist, so liefert der Aufruf `davisPutnam(clauses, literals)` eine Menge von Unit-Klauseln `r`, so dass jede Belegung \mathcal{I} , die alle Unit-Klauseln aus `r` erfüllt, auch die Menge `clauses` erfüllt. Falls die Menge `clauses` nicht erfüllbar ist, liefert der Aufruf `davisPutnam(Clauses, Literals)` als Ergebnis die Menge $\{\{\}\}$ zurück, denn die leere Klausel repräsentiert die unerfüllbare Formel \perp .

Sie fragen sich vielleicht, wozu wir in der Prozedur `davisPutnam` die Menge `literals` brauchen. Der Grund ist, dass wir uns bei den rekursiven Aufrufen merken müssen, welche Literale wir schon benutzt haben. Diese Literale sammeln wir in der Menge `literals`.

Die in Abbildung 5.9 gezeigte Implementierung funktioniert wie folgt:

1. In Zeile 2 reduzieren wir mit Hilfe der Methode `saturate` solange wie möglich die gegebene Klausel-Menge `clauses` mit Hilfe von Unit-Schnitten und entfernen alle Klauseln die durch Unit-Klauseln subsumiert werden.

2. Anschließend testen wir in Zeile 3, ob die so vereinfachte Klausel-Menge die leere Klausel enthält und geben in diesem Fall als Ergebnis die Menge $\{\{\}\}$ zurück.
3. Dann testen wir in Zeile 6, ob bereits alle Klauseln c aus der Menge **clauses** Unit-Klauseln sind. Wenn dies so ist, dann ist die Menge **clauses** trivial und wir geben diese Menge als Ergebnis zurück.
4. Andernfalls wählen wir in Zeile 9 ein Literal l aus der Menge **clauses**, dass wir noch nicht benutzt haben. Wir untersuchen dann in Zeile 10 rekursiv, ob die Menge

$$\mathbf{clauses} \cup \{\{1\}\}$$

lösbar ist. Dabei gibt es zwei Fälle:

- (a) Falls diese Menge lösbar ist, geben wir die Lösung dieser Menge als Ergebnis zurück.
- (b) Sonst prüfen wir rekursiv, ob die Menge

$$\mathbf{clauses} \cup \{\{-1\}\}$$

lösbar ist.

```

1  davisPutnam := procedure(clauses, literals) {
2      clauses := saturate(clauses);
3      if ({ } in clauses) {
4          return { { } };
5      }
6      if (forall (c in clauses | #c == 1)) {
7          return clauses;
8      }
9      l := selectLiteral(clauses, literals);
10     notL := negateLiteral(l);
11     r := davisPutnam(clauses + { {l} }, literals + { l, notL });
12     if (r != { { } }) {
13         return r;
14     }
15     return davisPutnam(clauses + { {notL} }, literals + { l, notL });
16 };

```

Abbildung 5.9: Die Prozedur **davisPutnam**.

Wir diskutieren nun die Hilfsprozeduren, die bei der Implementierung der Prozedur **davisPutnam** verwendet wurden. Als erstes besprechen wir die Funktion **saturate**. Diese Prozedur erhält eine Menge s von Klauseln als Eingabe und führt alle möglichen Unit-Schnitte und Unit-Subsumptionen durch. Die Prozedur **saturate** ist in Abbildung 5.10 auf Seite 109 gezeigt.

Die Implementierung von **saturate** funktioniert wie folgt:

1. Zunächst berechnen wir in Zeile 2 die Menge **units** aller Unit-Klauseln.
2. Dann initialisieren wir in Zeile 3 die Menge **used** als die leere Menge. In dieser Menge merken wir uns, welche Unit-Klauseln wir schon für Unit-Schnitte und Subsumptionen benutzt haben.
3. Solange die Menge **units** der Unit-Klauseln nicht leer ist, wählen wir in Zeile 5 eine beliebige Unit-Klausel **unit** aus.
4. In Zeile 6 fügen wir die Klausel **unit** zu der Menge **used** der benutzten Klausel hinzu.
5. In Zeile 7 extrahieren mit **arb** das Literal l der Klausel **unit**.

```

1  saturate := procedure(s) {
2      units := { k in s | #k == 1 };
3      used := {};
4      while (units != {}) {
5          unit := arb(units);
6          used := used + { unit };
7          l := arb(unit);
8          s := reduce(s, l);
9          units := { k in s | #k == 1 } - used;
10     }
11     return s;
12 };

```

Abbildung 5.10: Die Prozedur **saturate**.

6. In Zeile 8 wird die eigentliche Arbeit durch einen Aufruf der Prozedur **reduce** geleistet.
7. Dabei können jetzt neue Unit-Klauseln entstehen, die wir in Zeile 9 aufsammeln. Wir sammeln dort aber nur die Unit-Klauseln auf, die wir noch nicht benutzt haben.
8. Die Schleife in den Zeilen 4 – 10 wird nun solange durchlaufen, wie wir Unit-Klauseln finden, die wir noch nicht benutzt haben.
9. Am Ende geben wir die verbliebende Klauselmenge als Ergebnis zurück.

Die dabei verwendete Prozedur **reduce()** ist in Abbildung 5.11 gezeigt. Im vorigen Abschnitt hatten wir die Funktion $reduce(K, l)$, die eine Klausel-Menge K mit Hilfe des Literals l reduziert, als

$$reduce(K, l) = \{ k \setminus \{\bar{l}\} \mid k \in K \wedge \bar{l} \in k \} \cup \{ k \in K \mid \bar{l} \notin k \wedge l \notin k \} \cup \{\{l\}\}$$

definiert. Die Implementierung setzt diese Definition unmittelbar um.

```

1  reduce := procedure(s, l) {
2      notL := negateLiteral(l);
3      return { k - { notL } : k in s | notL in k }
4              + { k in s | !(notL in k) && !(l in k) }
5              + { {l} };
6  };

```

Abbildung 5.11: Die Prozedur **reduce**.

Die Implementierung des Algorithmus von Davis und Putnam benutzt außer den bisher diskutierten Prozeduren noch zwei weitere Hilfsprozeduren, deren Implementierung in Abbildung 5.12 auf Seite 110 gezeigt wird.

1. Die Prozedur **selectLiteral** wählt ein beliebiges Literal aus einer gegebenen Menge s von Klauseln aus, das außerdem nicht in der Menge **forbidden** von Literalen vorkommen darf, die bereits benutzt worden sind. Dazu werden alle Klauseln, die ja Mengen von Literalen sind, vereinigt. Von dieser Menge wird dann die Menge der bereits benutzten Literalen abgezogen und aus der resultierenden Menge wird mit Hilfe der Funktion **arb()** ein Literal ausgewählt.
2. Die Prozedur **negateLiteral** bildet die Negation \bar{l} eines gegebenen Literals l .

```

1  selectLiteral := procedure(s, forbidden) {
2      return arb(+/ s - forbidden);
3  };
4  negateLiteral := procedure(l) {
5      match (l) {
6          case !p : return p;
7          case  p : return !p;
8      }
9  };

```

Abbildung 5.12: Die Prozeduren `select` und `negateLiteral`.

Die oben dargestellte Version des Verfahrens von Davis und Putnam lässt sich in vielerlei Hinsicht verbessern. Aus Zeitgründen können wir auf solche Verbesserungen leider nicht weiter eingehen. Der interessierte Leser sei hier auf [MMZ⁺01] verwiesen.

Chaff: Engineering an Efficient SAT Solver

von M. Moskewicz, C. Madigan, Y. Zhao, L. Zhang, S. Malik

5.7 Das 8-Damen-Problem

In diesem Abschnitt zeigen wir, wie bestimmte kombinatorische Problem in aussagenlogische Probleme umformuliert werden können. Diese können dann anschließend mit dem Algorithmus von Davis und Putnam bzw. mit Verbesserungen dieses Algorithmus gelöst werden. Als konkretes Beispiel betrachten wir das 8-Damen-Problem. Dabei geht es darum, 8 Damen so auf einem Schach-Brett aufzustellen, dass keine Dame eine andere Dame schlagen kann. Beim Schach-Spiel kann eine Dame dann eine andere Figur schlagen, wenn diese Figur entweder

- in der selben Zeile,
- in der selben Spalte, oder
- in der selben Diagonale

wie die Dame steht. Abbildung 5.13 auf Seite 111 zeigt ein Schachbrett, in dem sich in der dritten Zeile in der vierten Spalte eine Dame befindet. Diese Dame kann auf alle die Felder ziehen, die mit Pfeilen markierte sind, und kann damit Figuren, die sich auf diesen Feldern befinden, schlagen.

Als erstes überlegen wir uns, wie wir ein Schach-Brett mit den darauf positionierten Damen aussagenlogisch repräsentieren können. Eine Möglichkeit besteht darin, für jedes Feld eine aussagenlogische Variable einzuführen. Diese Variable drückt aus, dass auf dem entsprechenden Feld eine Dame steht. Wir ordnen diesen Variablen wie folgt Namen zu: Die Variable, die das j -te Feld in der i -ten Zeile bezeichnet, stellen wir durch den Term

$$\text{Var}(i, j) \quad \text{mit } i, j \in \{1, \dots, 8\}$$

dar. Wir numerieren die Zeilen dabei von oben beginnend von 1 bis 8 durch, während die Spalten von links nach rechts numeriert werden. Abbildung 5.14 auf Seite 112 zeigt die Zuordnung der Variablen zu den Feldern.

Als nächstes überlegen wir uns, wie wir die einzelnen Bedingungen des 8-Damen-Problems als aussagenlogische Formeln kodieren können. Letztlich lassen sich alle Aussagen der Form

- “in einer Zeile steht höchstens eine Dame”,
- “in einer Spalte steht höchstens eine Dame”, oder
- “in einer Diagonale steht höchstens eine Dame”

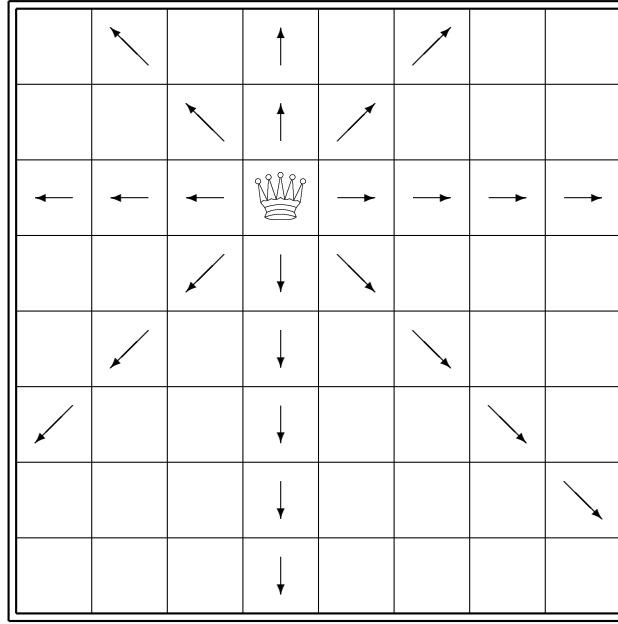


Abbildung 5.13: Das 8-Damen-Problem.

auf das selbe Grundmuster zurückführen: Ist eine Menge von aussagenlogischen Variablen

$$V = \{x_1, \dots, x_n\}$$

gegeben, so brauchen wir eine Formel die aussagt, dass **höchstens** eine der Variablen aus V den Wert **true** hat. Das ist aber gleichbedeutend damit, dass für jedes Paar $x_i, x_j \in V$ mit $x_i \neq x_j$ die folgende Formel gilt:

$$\neg(x_i \wedge x_j).$$

Diese Formel drückt aus, dass die Variablen x_i und x_j nicht gleichzeitig den Wert **true** annehmen. Nach den DeMorgan'schen Gesetzen gilt

$$\neg(x_i \wedge x_j) \leftrightarrow \neg x_i \vee \neg x_j$$

und die Klausel auf der rechten Seite dieser Äquivalenz schreibt sich in Mengen-Schreibweise als

$$\{\neg x_i, \neg x_j\}.$$

Die Formel, die für eine Variablen-Menge V ausdrückt, dass keine zwei verschiedenen Variablen gleichzeitig gesetzt sind, kann daher als Klausel-Menge als

$$\{ \{ \neg p, \neg q \} \mid p \in V \wedge q \in V \wedge p \neq q \}$$

geschrieben werden. Wir setzen diese Überlegungen in eine SETLX-Prozedur um. Die in Abbildung 5.15 gezeigte Prozedur `atMostOne()` bekommt als Eingabe eine Menge V von aussagenlogischen Variablen. Der Aufruf `atMostOne(V)` berechnet eine Menge von Klauseln. Diese Klauseln sind genau dann wahr, wenn höchstens eine der Variablen aus V den Wert **true** hat.

Mit Hilfe der Prozedur `atMostOne` können wir nun die Prozedur `atMostOneInRow` implementieren. Der Aufruf

$$\text{atMostOneInRow}(\text{row}, n)$$

berechnet für eine gegebene Zeile `row` bei einer Brettgröße von n eine Formel, die ausdrückt, dass in der Zeile `row` höchstens eine Dame steht. Abbildung 5.16 zeigt die Prozedur `atMostOneInRow()`: Wir sammeln alle Variablen der durch `row` spezifizierten Zeile in der Menge

$$\{\text{Var}(\text{row}, j) \mid j \in \{1, \dots, n\}\}$$

Var(1,1)	Var(1,2)	Var(1,3)	Var(1,4)	Var(1,5)	Var(1,6)	Var(1,7)	Var(1,8)
Var(2,1)	Var(2,2)	Var(2,3)	Var(2,4)	Var(2,5)	Var(2,6)	Var(2,7)	Var(2,8)
Var(3,1)	Var(3,2)	Var(3,3)	Var(3,4)	Var(3,5)	Var(3,6)	Var(3,7)	Var(3,8)
Var(4,1)	Var(4,2)	Var(4,3)	Var(4,4)	Var(4,5)	Var(4,6)	Var(4,7)	Var(4,8)
Var(5,1)	Var(5,2)	Var(5,3)	Var(5,4)	Var(5,5)	Var(5,6)	Var(5,7)	Var(5,8)
Var(6,1)	Var(6,2)	Var(6,3)	Var(6,4)	Var(6,5)	Var(6,6)	Var(6,7)	Var(6,8)
Var(7,1)	Var(7,2)	Var(7,3)	Var(7,4)	Var(7,5)	Var(7,6)	Var(7,7)	Var(7,8)
Var(8,1)	Var(8,2)	Var(8,3)	Var(8,4)	Var(8,5)	Var(8,6)	Var(8,7)	Var(8,8)

Abbildung 5.14: Zuordnung der Variablen.

```

1  atMostOne := procedure(v) {
2      return { { !p, !q } : p in v, q in v | p != q };
3  };

```

Abbildung 5.15: Die Prozedur `atMostOne`.

auf und rufen mit dieser Menge die Hilfs-Prozedur `atMostOne()` auf, die das Ergebnis als Menge von Klauseln liefert.

Als nächstes berechnen wir eine Formel die aussagt, dass mindestens eine Dame in einer gegebenen Spalte steht. Für die erste Spalte hätte diese Formel die Form

$$\text{Var}(1, 1) \vee \text{Var}(2, 1) \vee \text{Var}(3, 1) \vee \text{Var}(4, 1) \vee \text{Var}(5, 1) \vee \text{Var}(6, 1) \vee \text{Var}(7, 1) \vee \text{Var}(8, 1)$$

```

1  atMostOneInRow := procedure(row, n) {
2      return atMostOne({ Var(row, j) : j in [1 .. n] });
3  };

```

Abbildung 5.16: Die Prozedur `atMostOneInRow`.

und wenn allgemein eine Spalte c mit $c \in \{1, \dots, 8\}$ gegeben ist, lautet die Formel

$$\text{Var}(1, c) \vee \text{Var}(2, c) \vee \text{Var}(3, c) \vee \text{Var}(4, c) \vee \text{Var}(5, c) \vee \text{Var}(6, c) \vee \text{Var}(7, c) \vee \text{Var}(8, c).$$

Schreiben wir diese Formel in der Mengenschreibweise als Menge von Klauseln, so erhalten wir

$$\{\{\text{Var}(1, c), \text{Var}(2, c), \text{Var}(3, c), \text{Var}(4, c), \text{Var}(5, c), \text{Var}(6, c), \text{Var}(7, c), \text{Var}(8, c)\}\}.$$

Abbildung 5.17 zeigt eine SETLX-Prozedur, die für eine gegebene Spalte `column` die entsprechende Klausel-Menge berechnet. Der Schritt, von einer einzelnen Klausel zu einer Menge von Klauseln überzugehen ist notwendig, da unsere Implementierung des Algorithmus von Davis und Putnam ja mit einer Menge von Klauseln arbeitet.

```

1  oneInColumn := procedure(column, n) {
2      return { { Var(row, column) : row in { 1 .. n } } };
3  };

```

Abbildung 5.17: Die Prozedur `oneInColumn`.

An dieser Stelle erwarten Sie vielleicht, dass wir als noch Formeln angeben die ausdrücken, dass in einer gegebenen Spalte höchstens eine Dame steht und dass in jeder Reihe mindestens eine Dame steht. Solche Formeln sind aber unnötig, denn wenn wir wissen, dass in jeder Spalte mindestens eine Dame steht, so wissen wir bereits, dass auf dem Brett mindestens 8 Damen stehen. Wenn wir nun zusätzlich wissen, dass in jeder Zeile höchstens eine Dame steht, so ist automatisch klar, dass in jeder Zeile genau eine Dame stehen muß, denn sonst kommen wir insgesamt nicht auf 8 Damen. Weiter folgt aus der Tatsache, dass in jeder Spalte eine Dame steht und daraus, dass es insgesamt nicht mehr als 8 Damen sind, dass in jeder Spalte höchstens eine Dame stehen kann.

Als nächstes überlegen wir uns, wie wir die Variablen, die auf der selben Diagonale stehen, charakterisieren können. Es gibt grundsätzlich zwei verschiedene Arten von Diagonalen: absteigende Diagonalen und aufsteigende Diagonalen. Wir betrachten zunächst die aufsteigenden Diagonalen. Die längste aufsteigende Diagonale, wir sagen dazu auch *Hauptdiagonale*, besteht aus den Variablen

$$\text{Var}(8, 1), \text{Var}(7, 2), \text{Var}(6, 3), \text{Var}(5, 4), \text{Var}(4, 5), \text{Var}(3, 6), \text{Var}(2, 7), \text{Var}(1, 8).$$

Die Indizes i und j der Variablen $\text{Var}(i, j)$ erfüllen offenbar die Gleichung

$$i + j = 9.$$

Allgemein erfüllen die Indizes der Variablen einer aufsteigenden Diagonale die Gleichung

$$i + j = k,$$

wobei k einen Wert aus der Menge $\{3, \dots, 15\}$ annimmt. Diesen Wert k geben wir nun als Argument bei der Prozedur `atMostOneInUpperDiagonal` mit. Diese Prozedur ist in Abbildung 5.18 gezeigt.

Um zu sehen, wie die Variablen einer fallenden Diagonale charakterisiert werden können, betrachten wir die fallende Hauptdiagonale, die aus den Variablen

$$\text{Var}(1, 1), \text{Var}(2, 2), \text{Var}(3, 3), \text{Var}(4, 4), \text{Var}(5, 5), \text{Var}(6, 6), \text{Var}(7, 7), \text{Var}(8, 8)$$

besteht. Die Indizes i und j dieser Variablen erfüllen offenbar die Gleichung

$$i - j = 0.$$

```

1  atMostOneInUpperDiagonal := procedure(k, n) {
2      s := { Var(r, c) : c in [1..n], r in [1..n] | r + c == k };
3      return atMostOne(s);
4  };

```

Abbildung 5.18: Die Prozedur `atMostOneInUpperDiagonal`.

Allgemein erfüllen die Indizes der Variablen einer absteigenden Diagonale die Gleichung

$$i - j = k,$$

wobei k einen Wert aus der Menge $\{-6, \dots, 6\}$ annimmt. Diesen Wert k geben wir nun als Argument bei der Prozedur `atMostOneInLowerDiagonal` mit. Diese Prozedur ist in [Abbildung 5.19](#) gezeigt.

```

1  atMostOneInLowerDiagonal := procedure(k, n) {
2      s := { Var(r, c) : c in [1..n], r in [1..n] | r - c == k };
3      return atMostOne(s);
4  };

```

Abbildung 5.19: Die Prozedur `atMostOneInLowerDiagonal`.

Jetzt sind wir in der Lage, unsere Ergebnisse zusammen zu fassen: Wir können eine Menge von Klauseln konstruieren, die das 8-Damen-Problem vollständig beschreibt. [Abbildung 5.20](#) zeigt die Implementierung der Prozedur `allClauses`. Der Aufruf

`allClauses(n)`

rechnet für ein Schach-Brett der Größe n eine Menge von Klauseln aus, die genau dann erfüllt sind, wenn auf dem Schach-Brett

1. in jeder Zeile höchstens eine Dame steht (Zeile 2),
2. in jeder absteigenden Diagonale höchstens eine Dame steht (Zeile 3),
3. in jeder aufsteigenden Diagonale höchstens eine Dame steht (Zeile 4) und
4. in jeder Spalte mindestens eine Dame steht (Zeile 5).

Die Ausdrücke in den einzelnen Zeilen liefern Mengen, deren Elemente Klausel-Mengen sind. Was wir als Ergebnis brauchen ist aber eine Klausel-Menge und keine Menge von Klausel-Mengen. Daher bilden wir mit dem Operator “+” die Vereinigung dieser Mengen.

```

1  allClauses := procedure(n) {
2      return  +/ { atMostOneInRow(row, n)           : row in {1..n}           }
3              + +/ { atMostOneInLowerDiagonal(k, n) : k in {-(n-2) .. n-2} }
4              + +/ { atMostOneInUpperDiagonal(k, n) : k in {3 .. 2*n - 1} }
5              + +/ { oneInColumn(column, n)         : column in {1 .. n}     };
6  };

```

Abbildung 5.20: Die Prozedur `allClauses`.

Als letztes zeigen wir in [Abbildung 5.21](#) die Prozedur `solve`, mit der wir das 8-Damen-Problem lösen können. Hierbei ist `printBoard()` eine Prozedur, welche die Lösung in lesbarere Form als Schachbrett ausdrückt. Das

funktioniert allerdings nur, wenn ein Font verwendet wird, bei dem alle Zeichen die selbe Breite haben. Diese Prozedur ist der Vollständigkeit halber in Abbildung 5.22 gezeigt, wir wollen die Implementierung aber nicht weiter diskutieren. Das vollständige Programm finden Sie auf meiner Webseite unter dem Namen `queens.stlx`.

```
1  solve := procedure(n) {
2      clauses := allClauses(n);
3      solution := davisPutnam(clauses, {});
4      if (solution != { {} }) {
5          printBoard(solution, n);
6      } else {
7          print("The problem is not solvable for " + numberQueens + " queens!");
8          print("Try to increase the number of queens.");
9      }
10 };
```

Abbildung 5.21: Die Prozedur `solve`.

```
1  printBoard := procedure(i, n) {
2      if (i == { {} }) {
3          return;
4      }
5      print( "          " + ((8*n+1) * "-") );
6      for (row in [1..n]) {
7          line := "          |";
8          for (col in [1..n]) {
9              line += "          |";
10             }
11             print(line);
12             line := "          |";
13             for (col in [1..n]) {
14                 if ({ Var(row, col) } in i) {
15                     line += "    Q    |";
16                 } else {
17                     line += "          |";
18                 }
19             }
20             print(line);
21             line := "          |";
22             for (col in [1..n]) {
23                 line += "          |";
24             }
25             print(line);
26             print( "          " + ((8*n+1) * "-") );
27         }
28     };
```

Abbildung 5.22: Die Prozedur `printBoard()`.

Die durch den Aufruf `davisPutnam(Clauses, {})` berechnete Menge `solution` enthält für jede der Variablen $\text{Var}(i, j)$ entweder die Unit-Klausel $\{\text{Var}(i, j)\}$ (falls auf diesem Feld eine Dame steht) oder aber die Unit-Klausel $\{\neg \text{Var}(i, j)\}$ (falls das Feld leer bleibt). Eine graphische Darstellung des durch die berechnete Belegung dargestellten Schach-Bretts sehen Sie in Abbildung 5.23.

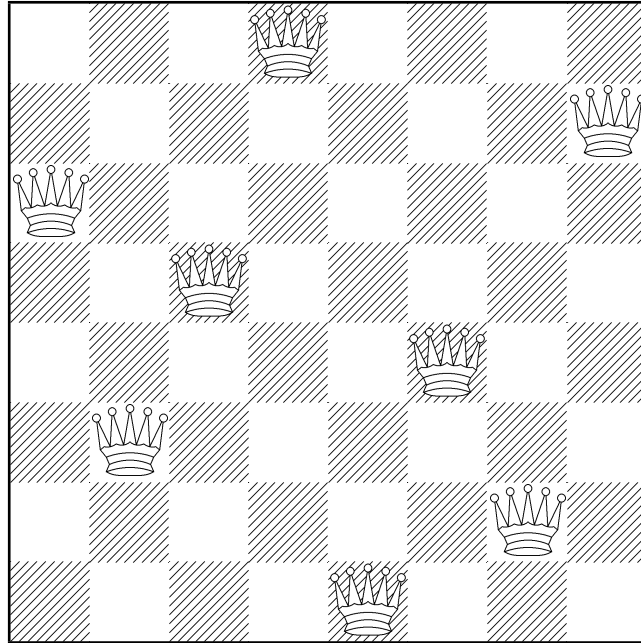


Abbildung 5.23: Eine Lösung des 8-Damen-Problems.

Das 8-Damen-Problem ist natürlich nur eine spielerische Anwendung der Aussagen-Logik. Trotzdem zeigt es die Leistungsfähigkeit des Algorithmus von Davis und Putnam sehr gut, denn die Menge der Klauseln, die von der Prozedur `allClauses` berechnet wird, füllt unformatiert fünf Bildschirm-Seiten, falls diese eine Breite von 80 Zeichen haben. In dieser Klausel-Menge kommen 64 verschiedene Variablen vor. Der Algorithmus von Davis und Putnam benötigt zur Berechnung einer Belegung, die diese Klauseln erfüllt, auf meinem iMac weniger als fünf Sekunden.

In der Praxis gibt es viele Probleme, die sich in ganz ähnlicher Weise auf die Lösung einer Menge von Klauseln zurückführen lassen. Dazu gehört zum Beispiel das Problem, einen Stundenplan zu erstellen, der gewissen Nebenbedingungen genügt. Verallgemeinerungen des Stundenplan-Problems werden in der Literatur als *Scheduling-Problemen* bezeichnet. Die effiziente Lösung solcher Probleme ist Gegenstand der aktuellen Forschung.

Kapitel 6

Prädikatenlogik

In der Aussagenlogik haben wir die Verknüpfung von elementaren Aussagen mit Junktoren untersucht. Die Prädikatenlogik untersucht zusätzlich auch die Struktur der Aussagen. Dazu werden in der Prädikatenlogik die folgenden zusätzlichen Begriffe eingeführt:

1. Als Bezeichnungen für Objekte werden *Terme* verwendet.
2. Diese Terme werden aus *Variablen* und *Funktions-Zeichen* zusammengesetzt:

$$\text{vater}(x), \quad \text{mutter}(\text{isaac}), \quad x + 7, \quad \dots$$

3. Verschiedene Objekte werden durch *Prädikats-Zeichen* in Relation gesetzt:

$$\text{istBruder}(\text{albert}, \text{vater}(\text{bruno})), \quad x + 7 < x \cdot 7, \quad n \in \mathbb{N}, \quad \dots$$

Die dabei entstehenden Formeln werden als *atomare* Formeln bezeichnet.

4. Atomare Formeln lassen sich durch aussagenlogische Junktoren verknüpfen:

$$x > 1 \rightarrow x + 7 < x \cdot 7$$

5. Schließlich werden *Quantoren* eingeführt, um zwischen *existentiell* und *universell* quantifizierten Variablen unterscheiden zu können:

$$\forall x \in \mathbb{R} : \exists n \in \mathbb{N} : x < n$$

Wir werden im nächsten Abschnitt die Syntax der prädikatenlogischen Formeln festlegen und uns dann im darauf folgenden Abschnitt mit der Semantik dieser Formeln beschäftigen.

6.1 Syntax der Prädikatenlogik

Zunächst definieren wir den Begriff der *Signatur*. Inhaltlich ist das nichts anderes als eine strukturierte Zusammenfassung von Variablen, Funktions- und Prädikats-Zeichen zusammen mit einer Spezifikation der Stelligkeit dieser Zeichen.

Definition 40 (Signatur) Eine *Signatur* ist ein 4-Tupel

$$\Sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, \text{arity} \rangle,$$

für das Folgendes gilt:

1. \mathcal{V} ist die Menge der Variablen.
2. \mathcal{F} ist die Menge der Funktions-Zeichen.
3. \mathcal{P} ist die Menge der Prädikats-Zeichen.
4. $arity$ ist eine Funktion, die jedem Funktions- und jedem Prädikats-Zeichen seine *Stelligkeit* zuordnet:

$$arity : \mathcal{F} \cup \mathcal{P} \rightarrow \mathbb{N}.$$

Wir sagen, dass das Funktions- oder Prädikats-Zeichen f ein n -stelliges Zeichen ist, falls $arity(f) = n$ gilt.

5. Da wir in der Lage sein müssen, Variablen, Funktions- und Prädikats-Zeichen unterscheiden zu können, vereinbaren wir, dass die Mengen \mathcal{V} , \mathcal{F} und \mathcal{P} paarweise disjunkt sein müssen:

$$\mathcal{V} \cap \mathcal{F} = \{\}, \quad \mathcal{V} \cap \mathcal{P} = \{\}, \quad \text{und} \quad \mathcal{F} \cap \mathcal{P} = \{\}.$$

Als Bezeichner für Objekte verwenden wir Ausdrücke, die aus Variablen und Funktions-Zeichen aufgebaut sind. Solche Ausdrücke nennen wir *Terme*. Formal werden diese wie folgt definiert.

Definition 41 (Terme) Ist $\Sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, arity \rangle$ eine Signatur, so definieren wir die Menge der Σ -Terme \mathcal{T}_Σ induktiv:

1. Für jede Variable $x \in \mathcal{V}$ gilt $x \in \mathcal{T}_\Sigma$.
2. Ist $f \in \mathcal{F}$ ein n -stelliges Funktions-Zeichen und sind $t_1, \dots, t_n \in \mathcal{T}_\Sigma$, so gilt auch

$$f(t_1, \dots, t_n) \in \mathcal{T}_\Sigma.$$

Falls $c \in \mathcal{F}$ ein 0-stelliges Funktions-Zeichen ist, lassen wir auch die Schreibweise c anstelle von $c()$ zu. In diesem Fall nennen wir c eine *Konstante*. \square

Beispiel: Es sei

1. $\mathcal{V} := \{x, y, z\}$ die Menge der Variablen,
2. $\mathcal{F} := \{0, 1, +, -, \cdot\}$ die Menge der Funktions-Zeichen,
3. $\mathcal{P} := \{=, \leq\}$ die Menge der Prädikats-Zeichen,
4. $arity := \{\langle 0, 0 \rangle, \langle 1, 0 \rangle, \langle +, 2 \rangle, \langle -, 2 \rangle, \langle \cdot, 2 \rangle, \langle =, 2 \rangle, \langle \leq, 2 \rangle\}$,
gibt die Stelligkeit der Funktions- und Prädikats-Zeichen an und
5. $\Sigma_{arith} := \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, arity \rangle$ sei eine Signatur.

Dann können wir wie folgt Σ_{arith} -Terme konstruieren:

1. $x, y, z \in \mathcal{T}_{\Sigma_{arith}}$,
denn alle Variablen sind auch Σ_{arith} -Terme.
2. $0, 1 \in \mathcal{T}_{\Sigma_{arith}}$,
denn 0 und 1 sind 0-stellige Funktions-Zeichen.
3. $+(0, x) \in \mathcal{T}_{\Sigma_{arith}}$,
denn es gilt $0 \in \mathcal{T}_{\Sigma_{arith}}$, $x \in \mathcal{T}_{\Sigma_{arith}}$ und $+$ ist ein 2-stelliges Funktions-Zeichen.
4. $\cdot(+(0, x), 1) \in \mathcal{T}_{\Sigma_{arith}}$,
denn $+(0, x) \in \mathcal{T}_{\Sigma_{arith}}$, $1 \in \mathcal{T}_{\Sigma_{arith}}$ und \cdot ist ein 2-stelliges Funktions-Zeichen.

Als nächstes definieren wir den Begriff der *atomaren Formeln*. Darunter verstehen wir solche Formeln, die man nicht in kleinere Formeln zerlegen kann, atomare Formeln enthalten also weder Junktoren noch Quantoren.

Definition 42 (Atomare Formeln) Gegeben sei eine Signatur $\Sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, \text{arity} \rangle$. Die Menge der atomaren Σ -Formeln \mathcal{A}_Σ wird wie folgt definiert: Ist $p \in \mathcal{P}$ ein n -stelliges Prädikats-Zeichen und sind n Σ -Terme t_1, \dots, t_n gegeben, so ist $p(t_1, \dots, t_n)$ eine atomare Σ -Formel:

$$p(t_1, \dots, t_n) \in \mathcal{A}_\Sigma.$$

Falls p ein 0-stelliges Prädikats-Zeichen ist, dann schreiben wir auch p anstelle von $p()$. In diesem Fall nennen wir p eine *Aussage-Variable*. \square

Setzen wir das obige Beispiel fort, so können wir sehen, dass

$$=(*+(0, x), 1), 0)$$

eine atomare Σ_{arith} -Formel ist. Beachten Sie, dass wir bisher noch nichts über den Wahrheitswert von solchen Formeln ausgesagt haben. Die Frage, wann eine Formel als wahr oder falsch gelten soll, wird erst im nächsten Abschnitt untersucht.

Bei der Definition der prädikatenlogischen Formeln ist es notwendig, zwischen sogenannten *gebundenen* und *freien* Variablen zu unterscheiden. Wir führen diese Begriffe zunächst informal mit Hilfe eines Beispiels aus der Analysis ein. Wir betrachten die folgende Identität:

$$\int_0^x y \cdot t \, dt = \frac{1}{2} x^2 \cdot y$$

In dieser Gleichung treten die Variablen x und y *frei* auf, während die Variable t durch das Integral *gebunden* wird. Damit meinen wir folgendes: Wir können in dieser Gleichung für x und y beliebige Werte einsetzen, ohne dass sich an der Gültigkeit der Formel etwas ändert. Setzen wir zum Beispiel für x den Wert 2 ein, so erhalten wir

$$\int_0^2 y \cdot t \, dt = \frac{1}{2} 2^2 \cdot y$$

und diese Identität ist ebenfalls gültig. Demgegenüber macht es keinen Sinn, wenn wir für die gebundene Variable t eine Zahl einsetzen würden. Die linke Seite der entstehenden Gleichung wäre einfach undefiniert. Wir können für t höchstens eine andere Variable einsetzen. Ersetzen wir die Variable t beispielsweise durch u , so erhalten wir

$$\int_0^x y \cdot u \, du = \frac{1}{2} x^2 \cdot y$$

und das ist die selbe Aussage wie oben. Das funktioniert allerdings nicht mit jeder Variablen. Setzen wir für t die Variable y ein, so erhalten wir

$$\int_0^x y \cdot y \, dy = \frac{1}{2} x^2 \cdot y.$$

Diese Aussage ist aber falsch! Das Problem liegt darin, dass bei der Ersetzung von t durch y die vorher freie Variable y gebunden wurde.

Ein ähnliches Problem erhalten wir, wenn wir für y beliebige Terme einsetzen. Solange diese Terme die Variable t nicht enthalten, geht alles gut. Setzen wir beispielsweise für y den Term x^2 ein, so erhalten wir

$$\int_0^x x^2 \cdot t \, dt = \frac{1}{2} x^2 \cdot x^2$$

und diese Formel ist gültig. Setzen wir allerdings für y den Term t^2 ein, so erhalten wir

$$\int_0^x t^2 \cdot t \, dt = \frac{1}{2} x^2 \cdot t^2$$

und diese Formel ist nicht mehr gültig.

In der Prädikatenlogik binden die Quantoren “ \forall ” (*für alle*) und “ \exists ” (*es gibt*) Variablen in ähnlicher Weise, wie der Integral-Operator “ $\int \cdot dt$ ” in der Analysis Variablen bindet. Die oben gemachten Ausführungen zeigen, dass es zwei verschiedene Arten von Variable gibt: *freie Variable* und *gebundene Variable*. Um diese Begriffe präzisieren zu können, definieren wir zunächst für einen Σ -Term t die Menge der in t enthaltenen Variablen.

Definition 43 ($\text{Var}(t)$) Ist t ein Σ -Term, mit $\Sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, \text{arity} \rangle$, so definieren wir die Menge $\text{Var}(t)$ der Variablen, die in t auftreten, durch Induktion nach dem Aufbau des Terms:

1. $\text{Var}(x) := \{x\}$ für alle $x \in \mathcal{V}$,
2. $\text{Var}(f(t_1, \dots, t_n)) := \text{Var}(t_1) \cup \dots \cup \text{Var}(t_n)$. □

Definition 44 (Σ -Formel, gebundene und freie Variablen)

Es sei $\Sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, \text{arity} \rangle$ eine Signatur. Die Menge der Σ -Formeln bezeichnen wir mit \mathbb{F}_Σ . Wir definieren diese Menge induktiv. Gleichzeitig definieren wir für jede Formel $F \in \mathbb{F}_\Sigma$ die Menge $BV(F)$ der in F *gebunden* auftretenden Variablen und die Menge $FV(F)$ der in F *frei* auftretenden Variablen.

1. Es gilt $\perp \in \mathbb{F}_\Sigma$ und $\top \in \mathbb{F}_\Sigma$ und wir definieren

$$FV(\perp) := FV(\top) := BV(\perp) := BV(\top) := \{\}.$$

2. Ist $F = p(t_1, \dots, t_n)$ eine atomare Σ -Formel, so gilt $F \in \mathbb{F}_\Sigma$. Weiter definieren wir:

- (a) $FV(p(t_1, \dots, t_n)) := \text{Var}(t_1) \cup \dots \cup \text{Var}(t_n)$.
- (b) $BV(p(t_1, \dots, t_n)) := \{\}$.

3. Ist $F \in \mathbb{F}_\Sigma$, so gilt $\neg F \in \mathbb{F}_\Sigma$. Weiter definieren wir:

- (a) $FV(\neg F) := FV(F)$.
- (b) $BV(\neg F) := BV(F)$.

4. Sind $F, G \in \mathbb{F}_\Sigma$ und gilt außerdem

$$FV(F) \cap BV(G) = \{\} \quad \text{und} \quad FV(G) \cap BV(F) = \{\},$$

so gilt auch

- (a) $(F \wedge G) \in \mathbb{F}_\Sigma$,
- (b) $(F \vee G) \in \mathbb{F}_\Sigma$,
- (c) $(F \rightarrow G) \in \mathbb{F}_\Sigma$,
- (d) $(F \leftrightarrow G) \in \mathbb{F}_\Sigma$.

Weiter definieren wir für alle Junktoren $\odot \in \{\wedge, \vee, \rightarrow, \leftrightarrow\}$:

- (a) $FV(F \odot G) := FV(F) \cup FV(G)$.
- (b) $BV(F \odot G) := BV(F) \cup BV(G)$.

5. Sei $x \in \mathcal{V}$ und $F \in \mathbb{F}_\Sigma$ mit $x \notin BV(F)$. Dann gilt:

- (a) $(\forall x: F) \in \mathbb{F}_\Sigma$.
- (b) $(\exists x: F) \in \mathbb{F}_\Sigma$.

Weiter definieren wir

- (a) $FV((\forall x: F)) := FV((\exists x: F)) := FV(F) \setminus \{x\}$.
- (b) $BV((\forall x: F)) := BV((\exists x: F)) := BV(F) \cup \{x\}$.

Ist die Signatur Σ aus dem Zusammenhang klar oder aber unwichtig, so schreiben wir auch \mathbb{F} statt \mathbb{F}_Σ und sprechen dann einfach von Formeln statt von Σ -Formeln. □

Bei der oben gegebenen Definition haben wir darauf geachtet, dass eine Variable nicht gleichzeitig frei und gebunden in einer Formel auftreten kann, denn durch eine leichte Induktion nach dem Aufbau der Formeln lässt sich zeigen, dass für alle $F \in \mathbb{F}_\Sigma$ folgendes gilt:

$$FV(F) \cap BV(F) = \{\}.$$

Beispiel: Setzen wir das oben begonnene Beispiel fort, so sehen wir, dass

$$(\exists x: \leq (+ (y, x), y))$$

eine Formel aus $\mathbb{F}_{\Sigma_{\text{arith}}}$ ist. Die Menge der gebundenen Variablen ist $\{x\}$, die Menge der freien Variablen ist $\{y\}$.

Wenn wir Formeln immer in dieser Form anschreiben würden, dann würde die Lesbarkeit unverhältnismäßig leiden. Zur Abkürzung vereinbaren wir, dass in der Prädikatenlogik die selben Regeln zur Klammer-Ersparnis gelten sollen, die wir schon in der Aussagenlogik verwendet haben. Zusätzlich werden gleiche Quantoren zusammengefasst: Beispielsweise schreiben wir

$$\forall x, y: p(x, y) \quad \text{statt} \quad \forall x: (\forall y: p(x, y)).$$

Darüber hinaus legen wir fest, dass Quantoren stärker binden als die aussagenlogischen Junktoren. Damit können wir

$$\forall x: p(x) \wedge G \quad \text{statt} \quad (\forall x: p(x)) \wedge G$$

schreiben. Außerdem vereinbaren wir, dass wir zweistellige Prädikats- und Funktions-Zeichen auch in Infix-Notation angeben dürfen. Um eine eindeutige Lesbarkeit zu erhalten, müssen wir dann gegebenenfalls Klammern setzen. Wir schreiben beispielsweise

$$\mathbf{n}_1 = \mathbf{n}_2 \quad \text{anstelle von} \quad = (\mathbf{n}_1, \mathbf{n}_2).$$

Die Formel $(\exists x: \leq (+ (y, x), y))$ wird dann lesbarer als

$$\exists x: y + x \leq y$$

geschrieben. Außerdem finden Sie in der Literatur häufig Ausdrücke der Form $\forall x \in M : F$ oder $\exists x \in M : F$. Hierbei handelt es sich um Abkürzungen, die wie folgt definiert sind:

$$\begin{aligned} (\forall x \in M : F) &\stackrel{\text{def}}{\iff} \forall x : (x \in M \rightarrow F), \\ (\exists x \in M : F) &\stackrel{\text{def}}{\iff} \exists x : (x \in M \wedge F). \end{aligned}$$

6.2 Semantik der Prädikatenlogik

Als nächstes legen wir die Bedeutung der Formeln fest. Dazu definieren wir mit Hilfe der Mengenlehre den Begriff einer Σ -Struktur. Eine solche Struktur legt fest, wie die Funktions- und Prädikats-Zeichen der Signatur Σ zu interpretieren sind.

Definition 45 (Struktur) Es sei eine Signatur

$$\Sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, \text{arity} \rangle.$$

gegeben. Eine Σ -Struktur \mathcal{S} ist ein Paar $\langle \mathcal{U}, \mathcal{J} \rangle$, so dass folgendes gilt:

1. \mathcal{U} ist eine nicht-leere Menge. Diese Menge nennen wir auch das *Universum* der Σ -Struktur. Dieses Universum enthält die Werte, die sich später bei der Auswertung der Terme ergeben werden.
2. \mathcal{J} ist die *Interpretation* der Funktions- und Prädikats-Zeichen. Formal definieren wir \mathcal{J} als eine Abbildung mit folgenden Eigenschaften:

- (a) Jedem Funktions-Zeichen $f \in \mathcal{F}$ mit $\text{arity}(f) = m$ wird eine m -stellige Funktion

$$f^{\mathcal{J}}: \mathcal{U} \times \dots \times \mathcal{U} \rightarrow \mathcal{U}$$

zugeordnet, die m -Tupel des Universums \mathcal{U} in das Universum \mathcal{U} abbildet.

- (b) Jedem Prädikats-Zeichen $p \in \mathcal{P}$ mit $\text{arity}(p) = n$ wird eine n -stellige Funktion

$$p^{\mathcal{J}} : \mathcal{U} \times \cdots \times \mathcal{U} \rightarrow \mathbb{B}$$

zugeordnet, die jedem n -Tupel des Universums \mathcal{U} einen Wahrheitswert aus der Menge $\mathbb{B} = \{\mathbf{true}, \mathbf{false}\}$ zuordnet.

- (c) Ist das Zeichen “=” ein Element der Menge der Prädikats-Zeichen \mathcal{P} , so gilt

$$=^{\mathcal{J}}(u, v) = \mathbf{true} \quad \text{g.d.w.} \quad u = v,$$

das Gleichheits-Zeichen wird also durch die identische Relation $\text{id}_{\mathcal{U}}$ interpretiert.

Beispiel: Wir geben ein Beispiel für eine Σ_{arith} -Struktur $\mathcal{S}_{\text{arith}} = \langle \mathcal{U}_{\text{arith}}, \mathcal{J}_{\text{arith}} \rangle$, indem wir definieren:

1. $\mathcal{U}_{\text{arith}} = \mathbb{N}$.
2. Die Abbildung $\mathcal{J}_{\text{arith}}$ legen wir dadurch fest, dass die Funktions-Zeichen $0, 1, +, -, \cdot$ durch die entsprechend benannten Funktionen auf der Menge \mathbb{N} der natürlichen Zahlen zu interpretieren sind.
Ebenso sollen die Prädikats-Zeichen $=$ und \leq durch die Gleichheits-Relation und die Kleiner-Gleich-Relation interpretiert werden.

Beispiel: Wir geben ein weiteres Beispiel. Die Signatur Σ_G der Gruppen-Theorie sei definiert als

$$\Sigma_G = \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, \text{arity} \rangle \quad \text{mit}$$

1. $\mathcal{V} := \{x, y, z\}$
2. $\mathcal{F} := \{1, *\}$
3. $\mathcal{P} := \{=\}$
4. $\text{arity} = \{\langle 1, 0 \rangle, \langle *, 2 \rangle, \langle =, 2 \rangle\}$

Dann können wir eine Σ_G Struktur $\mathcal{Z} = \langle \{a, b\}, \mathcal{J} \rangle$ definieren, indem wir die Interpretation \mathcal{J} wie folgt festlegen:

1. $1^{\mathcal{J}} := a$
2. $*^{\mathcal{J}} := \left\{ \langle \langle a, a \rangle, a \rangle, \langle \langle a, b \rangle, b \rangle, \langle \langle b, a \rangle, b \rangle, \langle \langle b, b \rangle, a \rangle \right\}$
3. $=^{\mathcal{J}}$ ist die Identität:
 $=^{\mathcal{J}} := \left\{ \langle \langle a, a \rangle, \mathbf{true} \rangle, \langle \langle a, b \rangle, \mathbf{false} \rangle, \langle \langle b, a \rangle, \mathbf{false} \rangle, \langle \langle b, b \rangle, \mathbf{true} \rangle \right\}$

Beachten Sie, dass wir bei der Interpretation des Gleichheits-Zeichens keinen Spielraum haben!

Falls wir Terme auswerten wollen, die Variablen enthalten, so müssen wir für diese Variablen irgendwelche Werte aus dem Universum einsetzen. Welche Werte wir einsetzen, kann durch eine *Variablen-Belegung* festgelegt werden. Diesen Begriff definieren wir nun.

Definition 46 (Variablen-Belegung) Es sei eine Signatur

$$\Sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, \text{arity} \rangle$$

gegeben. Weiter sei $\mathcal{S} = \langle \mathcal{U}, \mathcal{J} \rangle$ eine Σ -Struktur. Dann bezeichnen wir eine Abbildung

$$\mathcal{I} : \mathcal{V} \rightarrow \mathcal{U}$$

als eine \mathcal{S} -Variablen-Belegung.

Ist \mathcal{I} eine \mathcal{S} -Variablen-Belegung, $x \in \mathcal{V}$ und $c \in \mathcal{U}$, so bezeichnet $\mathcal{I}[x/c]$ die Variablen-Belegung, die der Variablen x den Wert c zuordnet und die ansonsten mit \mathcal{I} übereinstimmt:

$$\mathcal{I}[x/c](y) := \begin{cases} c & \text{falls } y = x; \\ \mathcal{I}(y) & \text{sonst.} \end{cases}$$

□

Definition 47 (Semantik der Terme) Ist $\mathcal{S} = \langle \mathcal{U}, \mathcal{J} \rangle$ eine Σ -Struktur und \mathcal{I} eine \mathcal{S} -Variablen-Belegung, so definieren wir für jeden Term t den Wert $\mathcal{S}(\mathcal{I}, t)$ durch Induktion über den Aufbau von t :

1. Für Variablen $x \in \mathcal{V}$ definieren wir:

$$\mathcal{S}(\mathcal{I}, x) := \mathcal{I}(x).$$

2. Für Σ -Terme der Form $f(t_1, \dots, t_n)$ definieren wir

$$\mathcal{S}(\mathcal{I}, f(t_1, \dots, t_n)) := f^{\mathcal{J}}(\mathcal{S}(\mathcal{I}, t_1), \dots, \mathcal{S}(\mathcal{I}, t_n)).$$

□

Beispiel: Mit der oben definierten Σ_{arith} -Struktur $\mathcal{S}_{\text{arith}}$ definieren wir eine $\mathcal{S}_{\text{arith}}$ -Variablen-Belegung \mathcal{I} durch

$$\mathcal{I} := \{ \langle x, 0 \rangle, \langle y, 7 \rangle, \langle z, 42 \rangle \},$$

es gilt also

$$\mathcal{I}(x) := 0, \quad \mathcal{I}(y) := 7, \quad \text{und} \quad \mathcal{I}(z) := 42.$$

Dann gilt offenbar

$$\mathcal{S}(\mathcal{I}, x + y) = 7.$$

Definition 48 (Semantik der atomaren Σ -Formeln) Ist \mathcal{S} eine Σ -Struktur und \mathcal{I} eine \mathcal{S} -Variablen-Belegung, so definieren wir für jede atomare Σ -Formel $p(t_1, \dots, t_n)$ den Wert $\mathcal{S}(\mathcal{I}, p(t_1, \dots, t_n))$ wie folgt:

$$\mathcal{S}(\mathcal{I}, p(t_1, \dots, t_n)) := p^{\mathcal{J}}(\mathcal{S}(\mathcal{I}, t_1), \dots, \mathcal{S}(\mathcal{I}, t_n)).$$

□

Beispiel: In Fortführung des obigen Beispiels gilt:

$$\mathcal{S}(\mathcal{I}, z \leq x + y) = \text{false}.$$

Um die Semantik beliebiger Σ -Formeln definieren zu können, nehmen wir an, dass wir, genau wie in der Aussagenlogik, die folgenden Funktionen zur Verfügung haben:

1. $\neg: \mathbb{B} \rightarrow \mathbb{B}$,
2. $\vee: \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$,
3. $\wedge: \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$,
4. $\rightarrow: \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$,
5. $\leftrightarrow: \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$.

Die Semantik dieser Funktionen hatten wir durch die Tabelle in Abbildung 5.1 auf Seite 81 gegeben.

Definition 49 (Semantik der Σ -Formeln) Ist \mathcal{S} eine Σ -Struktur und \mathcal{I} eine \mathcal{S} -Variablen-Belegung, so definieren wir für jede Σ -Formel F den Wert $\mathcal{S}(\mathcal{I}, F)$ durch Induktion über den Aufbau von F :

1. $\mathcal{S}(\mathcal{I}, \top) := \text{true}$ und $\mathcal{S}(\mathcal{I}, \perp) := \text{false}$.
2. $\mathcal{S}(\mathcal{I}, \neg F) := \neg(\mathcal{S}(\mathcal{I}, F))$.
3. $\mathcal{S}(\mathcal{I}, F \wedge G) := \wedge(\mathcal{S}(\mathcal{I}, F), \mathcal{S}(\mathcal{I}, G))$.
4. $\mathcal{S}(\mathcal{I}, F \vee G) := \vee(\mathcal{S}(\mathcal{I}, F), \mathcal{S}(\mathcal{I}, G))$.
5. $\mathcal{S}(\mathcal{I}, F \rightarrow G) := \rightarrow(\mathcal{S}(\mathcal{I}, F), \mathcal{S}(\mathcal{I}, G))$.
6. $\mathcal{S}(\mathcal{I}, F \leftrightarrow G) := \leftrightarrow(\mathcal{S}(\mathcal{I}, F), \mathcal{S}(\mathcal{I}, G))$.
7. $\mathcal{S}(\mathcal{I}, \forall x: F) := \begin{cases} \text{true} & \text{falls } \mathcal{S}(\mathcal{I}[x/c], F) = \text{true} \text{ für alle } c \in \mathcal{U} \text{ gilt;} \\ \text{false} & \text{sonst.} \end{cases}$

$$8. \mathcal{S}(\mathcal{I}, \exists x: F) := \begin{cases} \text{true} & \text{falls } \mathcal{S}(\mathcal{I}[x/c], F) = \text{true} \text{ für ein } c \in \mathcal{U} \text{ gilt;} \\ \text{false} & \text{sonst.} \end{cases} \quad \square$$

Beispiel: In Fortführung des obigen Beispiels gilt

$$\mathcal{S}(\mathcal{I}, \forall x : x * 0 < 1) = \text{true}.$$

Definition 50 (Allgemeingültig) Ist F eine Σ -Formel, so dass für jede Σ -Struktur \mathcal{S} und für jede \mathcal{S} -Variablen-Belegung \mathcal{I}

$$\mathcal{S}(\mathcal{I}, F) = \text{true}$$

gilt, so bezeichnen wir F als *allgemeingültig*. In diesem Fall schreiben wir

$$\models F. \quad \square$$

Ist F eine Formel für die $FV(F) = \{\}$ ist, dann hängt der Wert $\mathcal{S}(\mathcal{I}, F)$ offenbar gar nicht von der Interpretation \mathcal{I} ab. Solche Formeln bezeichnen wir auch als *geschlossene* Formeln. In diesem Fall schreiben wir kürzer $\mathcal{S}(F)$ an Stelle von $\mathcal{S}(\mathcal{I}, F)$. Gilt dann zusätzlich $\mathcal{S}(F) = \text{true}$, so sagen wir auch dass \mathcal{S} ein *Modell* von F ist. Wir schreiben dann

$$\mathcal{S} \models F.$$

Die Definition der Begriffe “*erfüllbar*” und “*äquivalent*” lassen sich nun aus der Aussagenlogik übertragen. Um unnötigen Ballast in den Definitionen zu vermeiden, nehmen wir im folgenden immer eine feste Signatur Σ als gegeben an. Dadurch können wir in den folgenden Definitionen von Termen, Formeln, Strukturen, etc. sprechen und meinen damit Σ -Terme, Σ -Formeln und Σ -Strukturen.

Definition 51 (Äquivalent) Zwei Formeln F und G heißen *äquivalent* g.d.w. gilt

$$\models F \leftrightarrow G \quad \square$$

Alle aussagenlogischen Äquivalenzen sind auch prädikatenlogische Äquivalenzen.

Definition 52 (Erfüllbar) Eine Menge $M \subseteq \mathbb{F}_\Sigma$ ist genau dann *erfüllbar*, wenn es eine Struktur \mathcal{S} und eine Variablen-Belegung \mathcal{I} gibt, so dass

$$\forall m \in M : \mathcal{S}(\mathcal{I}, m) = \text{true}$$

gilt. Andernfalls heißt M *unerfüllbar* oder auch *widersprüchlich*. Wir schreiben dafür auch

$$M \models \perp \quad \square$$

Unser Ziel ist es, ein Verfahren anzugeben, mit dem wir in der Lage sind zu überprüfen, ob eine Menge M von Formeln *widersprüchlich* ist, ob also $M \models \perp$ gilt. Es zeigt sich, dass dies im Allgemeinen nicht möglich ist, die Frage, ob $M \models \perp$ gilt, ist unentscheidbar. Ein Beweis dieser Tatsache geht allerdings über den Rahmen dieser Vorlesung hinaus. Dem gegenüber ist es möglich, ähnlich wie in der Aussagenlogik einen *Kalkül* \vdash anzugeben, so dass gilt

$$M \vdash \perp \quad \text{g.d.w.} \quad M \models \perp.$$

Ein solcher Kalkül kann dann zur Implementierung eines *Semi-Entscheidungs-Verfahrens* benutzt werden: Um zu überprüfen, ob $M \models \perp$ gilt, versuchen wir, aus der Menge M die Formel \perp herzuleiten. Falls wir dabei systematisch vorgehen, indem wir alle möglichen Beweise durchprobieren, so werden wir, falls tatsächlich $M \models \perp$ gilt, auch irgendwann einen Beweis finden, der $M \vdash \perp$ zeigt. Wenn allerdings der Fall

$$M \not\models \perp$$

vorliegt, so werden wir dies im allgemeinen nicht feststellen können, denn die Menge aller Beweise ist unendlich groß und wir können nie alle Beweise ausprobieren. Wir können lediglich sicherstellen, dass wir jeden Beweis irgendwann versuchen. Wenn es aber keinen Beweis gibt, so können wir das nie sicher sagen, denn zu jedem festen Zeitpunkt haben wir ja immer nur einen Teil der in Frage kommenden Schlüsse ausprobiert.

Die Situation ist ähnlich der, wie bei der Überprüfung bestimmter zahlentheoretischer Fragen. Wir betrachten dazu ein konkretes Beispiel: Eine Zahl n heißt *perfekt*, wenn die Summe aller echten Teiler von n wieder die

Zahl n ergibt. Beispielsweise ist die Zahl 6 perfekt, denn die Menge der echten Teiler von 6 ist $\{1, 2, 3\}$ und es gilt

$$1 + 2 + 3 = 6.$$

Bisher sind alle bekannten perfekten Zahlen durch 2 teilbar. Die Frage, ob es auch ungerade Zahlen gibt, die perfekt sind, ist ein offenes mathematisches Problem. Um dieses Problem zu lösen könnten wir eine Programm schreiben, dass der Reihe nach für alle ungerade Zahlen überprüft, ob die Zahl perfekt ist. Abbildung 6.1 auf Seite 125 zeigt ein solches Programm. Wenn es eine ungerade perfekte Zahl gibt, dann wird dieses Programm diese Zahl auch irgendwann finden. Wenn es aber keine ungerade perfekte Zahl gibt, dann wird das Programm bis zum St. Nimmerleinstag rechnen und wir werden nie mit Sicherheit wissen, dass es keine ungeraden perfekten Zahlen gibt.

```

1  perfect := procedure(n) {
2      return +/ { x : x in {1 .. n-1} | n % x == 0 } == n;
3  };
4  findPerfect := procedure() {
5      n := 1;
6      while (true) {
7          if (perfect(n)) {
8              if (n % 2 == 0) {
9                  print(n);
10             } else {
11                 print("Heureka: Odd perfect number $n$ found!");
12             }
13         }
14         n := n + 1;
15     }
16 };
17 findPerfect();

```

Abbildung 6.1: Suche nach einer ungeraden perfekten Zahl.

In den nächsten Abschnitten gehen wir daran, den oben erwähnten Kalkül \vdash zu definieren. Es zeigt sich, dass die Arbeit wesentlich einfacher wird, wenn wir uns auf bestimmte Formeln, sogenannte *Klauseln*, beschränken. Wir zeigen daher zunächst im nächsten Abschnitt, dass jede Formel-Menge M so in eine Menge von Klauseln K transformiert werden kann, dass M genau dann erfüllbar ist, wenn K erfüllbar ist. Daher ist die Beschränkung auf Klauseln keine echte Einschränkung.

6.2.1 Implementierung prädikatenlogischer Strukturen in SetIX

Der im letzten Abschnitt präsentierte Begriff einer prädikatenlogischen Struktur erscheint zunächst sehr abstrakt. Wir wollen in diesem Abschnitt zeigen, dass sich dieser Begriff in einfacher Weise in SETLX implementieren lässt. Dadurch gelingt es, diesen Begriff zu veranschaulichen. Als konkretes Beispiel wollen wir Strukturen zu Gruppen-Theorie betrachten. Die Signatur Σ_G der Gruppen-Theorie war im letzten Abschnitt durch die Definition

$$\Sigma_G = \langle \{x, y, z\}, \{1, *\}, \{=\}, \{\langle 1, 0 \rangle, \langle *, 2 \rangle, \langle =, 2 \rangle\} \rangle$$

gegeben worden. Hierbei ist also “1” ein 0-stelliges Funktions-Zeichen, “*” ist eine 2-stellige Funktions-Zeichen und “=” ist ein 2-stelliges Prädikats-Zeichen. Wir hatten bereits eine Struktur \mathcal{S} angegeben, deren Universum aus der Menge $\{a, b\}$ besteht. In SetIX können wir diese Struktur durch den in Abbildung 6.2 gezeigten Code implementieren.

```

1  a := "a";
2  b := "b";
3  u := { a, b }; // the universe
4  product := { [ [ a, a ], a ], [ [ a, b ], b ], [ [ b, a ], b ], [ [ b, b ], a ] };
5  equals := { [ x, y ] : x in u, y in u | x == y };
6  j := { [ "E", a ], [ "^product", product ], [ "^equals", equals ] };
7  s := [ u, j ];
8  i := { [ "x", a ], [ "y", b ], [ "z", a ] };

```

Abbildung 6.2: Implementierung einer Struktur zur Gruppen-Theorie

1. Zur Abkürzung haben wir in den Zeile 1 und 2 die Variablen a und b als die Strings "a" und "b" definiert. Dadurch können wir weiter unten die Interpretation des Funktions-Zeichens "*" kürzer angeben.

2. Das in Zeile 3 definierte Universum u besteht aus den beiden Strings "a" und "b".

3. In Zeile 4 definieren wir eine Funktion **product** als binäre Relation. Für die so definierte Funktion gilt

$$\begin{aligned} \text{product}(\langle \text{"a"}, \text{"a"} \rangle) &= \text{"a"}, & \text{product}(\langle \text{"a"}, \text{"b"} \rangle) &= \text{"b"}, \\ \text{product}(\langle \text{"b"}, \text{"a"} \rangle) &= \text{"b"}, & \text{product}(\langle \text{"b"}, \text{"b"} \rangle) &= \text{"a"}. \end{aligned}$$

Diese Funktion verwenden wir später als die Interpretation $*^{\mathcal{J}}$ des Funktions-Zeichens "*".

4. Ebenso haben wir in Zeile 5 die Interpretation $=^{\mathcal{J}}$ des Prädikats-Zeichens "=" als die binäre Relation **equals** dargestellt.

5. In Zeile 6 fassen wir die einzelnen Interpretationen zu der Relation j zusammen, so dass für ein Funktions-Zeichen f die Interpretation $f^{\mathcal{J}}$ durch den Wert $j(f)$ gegeben ist.

Da wir später den in SETLX eingebauten Parser verwenden werden, stellen wir den Operator "*" durch das Funktions-Zeichen "^product" dar und das Prädikats-Zeichen "=" wird durch das Zeichen "^equals" dargestellt, denn dies sind die Namen, die von SETLX intern benutzt werden. Das neutrale Element "1" stellen wir durch das Funktions-Zeichen "E" dar, so dass später der Ausdruck "1" durch den Term "E()" repräsentiert wird.

6. Die Interpretation j wird dann in Zeile 7 mit dem Universum u zu der Struktur s zusammen gefasst.

7. Schließlich zeigt Zeile 8, dass eine Variablen-Belegung ebenfalls als Relation dargestellt werden kann. Die erste Komponente der Paare, aus denen diese Relation besteht, sind die Variablen. Die zweite Komponente ist ein Wert aus dem Universum.

Als nächstes überlegen wir uns, wie wir prädikatenlogische Formeln in einer solchen Struktur auswerten können. Abbildung 6.3 zeigt die Implementierung der Prozedur $evalFormula(f, S, I)$, der als Argumente eine prädikatenlogische Formel f , eine Struktur S und eine Variablen-Belegung I übergeben werden. Die Formel wird dabei als Term dargestellt, ganz ähnlich, wie wir das bei der Implementierung der Aussagenlogik schon praktiziert haben. Beispielsweise können wir die Formel

$$\forall x : \forall y : x * y = y * x$$

durch den Term

```

^forall(^variable("x"), ^forall(^variable("y"),
    ^equals(^product(^variable("x"), ^variable("y")),
        ^product(^variable("y"), ^variable("x"))
    )))

```

```

1 evalFormula := procedure(f, s, i) {
2     u := s[1];
3     match (f) {
4         case true      : return true;
5         case false     : return false;
6         case !g        : return !evalFormula(g, s, i);
7         case g && h     : return evalFormula(g, s, i) && evalFormula(h, s, i);
8         case g || h     : return evalFormula(g, s, i) || evalFormula(h, s, i);
9         case g => h     : return evalFormula(g, s, i) => evalFormula(h, s, i);
10        case g <==> h : return evalFormula(g, s, i) == evalFormula(h, s, i);
11        case forall (x in _ | g) :
12            return forall (c in u | evalFormula(g, s, modify(i, x, c)));
13        case exists (x in _ | g) :
14            return exists (c in u | evalFormula(g, s, modify(i, x, c)));
15        default : return evalAtomic(f, s, i); // atomic formula
16    }
17 };

```

Abbildung 6.3: Auswertung prädikatenlogischer Formeln

darstellen und dass ist im Wesentlichen auch die Struktur, die erzeugt wird, wenn wir den String

“forall (x in u | exists (y in u | x * y == E()))”

mit Hilfe der in SETLX vordefinierten Funktion `parse` in einen Term umwandeln.

Bemerkung: An dieser Stelle wundern Sie sich vermutlich, warum wir oben “x in _” und “y in _” schreiben, denn wir wollen die Variablen x und y eigentlich ja gar nicht einschränken. Der Grund ist, dass die Syntax von SETLX nur solche Quantoren erlaubt, in denen die Variablen auf eine Menge eingeschränkt sind. Daher sind wir gezwungen, bei der Verwendung von Quantoren die Variablen syntaktisch einzuschränken. Wir schreiben deswegen

forall (x in _ | g) bzw. exists (x in _ | g)

an Stelle von

forall (x | g) bzw. exists (x | g).

Da wir x hier nicht wirklich einschränken wollen, schreiben wir “x in _” und benutzen “_” als sogenannte *anonyme Variable*. ◇

Die Auswertung einer prädikatenlogischen Formel ist nun analog zu der in Abbildung 5.1 auf Seite 85 gezeigten Auswertung aussagenlogischer Formeln. Neu ist nur die Behandlung der Quantoren. In den Zeilen 11 und 12 behandeln wir die Auswertung allquantifizierter Formeln. Ist f eine Formel der Form $\forall y \in u: h$, so wird die Formel f durch den Term

$$f = \text{forall}(y, u, h)$$

dargestellt. Das Muster

forall (x in _ | g)

bindet daher x an die tatsächlich auftretende Variable y und g an die Teilformel h . Die Auswertung von $\forall x: g$ geschieht nach der Formel

$$\mathcal{S}(\mathcal{I}, \forall x: g) := \begin{cases} \text{true} & \text{falls } \mathcal{S}(\mathcal{I}[x/c], g) = \text{true} \text{ für alle } c \in \mathcal{U} \text{ gilt;} \\ \text{false} & \text{sonst.} \end{cases}$$

Um die Auswertung implementieren zu können, verwenden wir eine Prozedur *modify()*, welche die Variablen-Belegung *i* an der Stelle *x* zu *c* abändert, es gilt also

$$\text{modify}(\mathcal{I}, x, c) = \mathcal{I}[x/c].$$

Die Implementierung dieser Prozedur wird später in Abbildung 6.4 gezeigt. Bei der Auswertung eines All-Quantors können wir ausnutzen, dass die Sprache SETLX selber den Quantor **forall** unterstützt. Wir können also direkt testen, ob die Formel für alle möglichen Werte *c*, die wir für die Variable *x* einsetzen können, richtig ist. Die Auswertung eines Existenz-Quantors ist analog zur Auswertung eines All-Quantors.

```

1  evalAtomic := procedure(a, s, i) {
2      j := s[2];
3      p := fct(a); // predicate symbol
4      pJ := j[p];
5      argList := args(a);
6      argsVal := evalTermList(argList, s, i);
7      return argsVal in pJ;
8  };
9  evalTerm := procedure(t, s, i) {
10     if (fct(t) == "^variable") {
11         varName := args(t)[1];
12         return i[varName];
13     }
14     j := s[2];
15     f := fct(t); // function symbol
16     fJ := j[f];
17     argList := args(t);
18     argsVal := evalTermList(argList, s, i);
19     if (#argsVal > 0) {
20         result := fJ[argsVal];
21     } else {
22         result := fJ; // t is a constant
23     }
24     return result;
25 };
26 evalTermList := procedure(tl, s, i) {
27     return [ evalTerm(t, s, i) : t in tl ];
28 };
29 modify := procedure(i, v, c) {
30     x := args(v)[1]; // v = ^variable(x)
31     i[x] := c;
32     return i;
33 };

```

Abbildung 6.4: Auswertung von Termen und atomaren Formeln.

Abbildung 6.4 zeigt die Auswertung atomarer Formeln und prädikatenlogischer Terme. Um eine atomare Formel der Form

$$a = P(t_1, \dots, t_n)$$

auszuwerten, verschaffen wir uns in Zeile 4 zunächst die dem Prädikats-Zeichen *P* in der Struktur *S* zugeordnete Menge *pJ*. Anschließend werten wir die Argumente t_1, \dots, t_n aus und überprüfen dann, ob das Ergebnis dieser Auswertung tatsächlich ein Element der Menge *pJ* ist.

Die Prozedur `evalTerm()` arbeitet wie folgt: Das erste Argument t der Prozedur `evalTerm($t, \mathcal{S}, \mathcal{I}$)` ist der auszuwertende Term. Das zweite Argument \mathcal{S} ist eine prädikatenlogische Struktur und das dritte Argument \mathcal{I} ist eine Variablen-Belegung.

1. Falls t eine Variable ist, so geben wir in Zeile 12 einfach den Wert zurück, der in der Variablen-Belegung \mathcal{I} für diese Variable eingetragen ist. Die Variablen-Belegung wird dabei durch eine zweistellige Relation dargestellt, die wir als Funktion benutzen.
2. Falls der auszuwertende Term t die Form

$$t = F(t_1, \dots, t_n)$$

hat, werden in Zeile 18 zunächst rekursiv die Subterme t_1, \dots, t_n ausgewertet. Anschließend wird die Interpretation $F^{\mathcal{J}}$ des Funktions-Zeichens F herangezogen, um die Funktion $F^{\mathcal{J}}$ für die gegebenen Argumente auszuwerten, wobei in Zeile 20 der Fall betrachtet wird, dass tatsächlich Argumente vorhanden sind, während in Zeile 22 der Fall behandelt wird, dass es sich bei dem Funktions-Zeichen F um eine Konstante handelt, deren Wert dann unmittelbar durch $F^{\mathcal{J}}$ gegeben ist.

Die Implementierung der Prozedur `evalTermList()` wendet die Funktion `evalTerm()` auf alle Terme der gegebenen Liste an. Bei der Implementierung der in Zeile 31 gezeigten Prozedur `modify(I, x, c)`, die als Ergebnis die Variablen-Belegung $\mathcal{I}[x/c]$ berechnet, nutzen wir aus, dass wir bei einer Funktion, die als binäre Relation gespeichert ist, den Wert, der in dieser Relation für ein Argument x eingetragen ist, durch eine Zuweisung der Form $\mathcal{I}(x) := c$ abändern können.

```

1 g1 := parse("forall (x in u | x * E() == x)");
2 g2 := parse("forall (x in u | exists (y in u | x * y == E()))");
3 g3 := parse("forall (x in u | forall (y in u | forall (z in u | (x*y)*z == x*(y*z) )))");
4 gt := { g1, g2, g3 };
5
6 print("checking group theory in the structure ", s);
7 for (f in gt) {
8     print( "checking ", f, ": ", evalFormula(f, s, i) );
9 }

```

Abbildung 6.5: Axiome der Gruppen-Theorie

Wir zeigen nun, wie sich die in Abbildung 6.3 gezeigte Funktion `evalFormula($f, \mathcal{S}, \mathcal{I}$)` benutzen lässt um zu überprüfen, ob die in Abbildung 6.2 gezeigte Struktur die Axiome der *Gruppen-Theorie* erfüllt. Die Axiome der Gruppen-Theorie sind wie folgt:

1. Die Konstante 1 ist das rechts-neutrale Element der Multiplikation:

$$\forall x: x * 1 = x.$$

2. Für jedes Element x gibt es ein rechts-inverses Element y , dass mit dem Element x multipliziert die 1 ergibt:

$$\forall x: \exists y: x * y = 1.$$

3. Es gilt das Assoziativ-Gesetz:

$$\forall x: \forall y: \forall z: (x * y) * z = x * (y * z).$$

Diese Axiome sind in den Zeilen 1 bis 3 der Abbildung 6.5 wiedergegeben, wobei wir die “1” durch das Funktions-Zeichen “E” dargestellt haben. Die Schleife in den Zeilen 7 bis 9 überprüft schließlich, ob die Formeln in der oben definierten Struktur erfüllt sind.

Bemerkung: Mit dem oben vorgestellten Programm können wir überprüfen, ob eine prädikatenlogische Formel in einer vorgegebenen endlichen Struktur erfüllt ist. Wir können damit allerdings nicht überprüfen, ob eine Formel allgemeingültig ist, denn einerseits können wir das Programm nicht anwenden, wenn die Strukturen ein unendliches Universum haben, andererseits ist selbst die Zahl der verschiedenen endlichen Strukturen, die wir ausprobieren müßten, unendlich groß.

Aufgabe 5:

1. Zeigen Sie, dass die Formel

$$\forall x : \exists y : p(x, y) \rightarrow \exists y : \forall x : p(x, y)$$

nicht allgemeingültig ist, indem Sie eine geeignete prädikatenlogische Struktur \mathcal{S} implementieren, in der diese Formel falsch ist.

2. Entwickeln Sie ein SETLX-Programm, dass die obige Formel in allen Strukturen ausprobiert, in denen das Universum aus einer vorgegebenen Zahl n verschiedener Elemente besteht und testen Sie Ihr Programm für $n = 2$.
3. Überlegen Sie, wie viele verschiedene Strukturen mit n Elementen es für die obige Formel gibt.
4. Geben Sie eine erfüllbare prädikatenlogische Formel F an, die in einer prädikatenlogischen Struktur $\mathcal{S} = \langle \mathcal{U}, \mathcal{J} \rangle$ immer falsch ist, wenn das Universum \mathcal{U} endlich ist. \diamond

6.3 Normalformen für prädikatenlogische Formeln

In diesem Abschnitt werden wir verschiedenen Möglichkeiten zur Umformung prädikatenlogischer Formeln kennenlernen. Zunächst geben wir einige Äquivalenzen an, mit deren Hilfe Quantoren manipuliert werden können.

Satz 53 Es gelten die folgenden Äquivalenzen:

1. $\models \neg(\forall x: f) \leftrightarrow (\exists x: \neg f)$
2. $\models \neg(\exists x: f) \leftrightarrow (\forall x: \neg f)$
3. $\models (\forall x: f) \wedge (\forall x: g) \leftrightarrow (\forall x: f \wedge g)$
4. $\models (\exists x: f) \vee (\exists x: g) \leftrightarrow (\exists x: f \vee g)$
5. $\models (\forall x: \forall y: f) \leftrightarrow (\forall y: \forall x: f)$
6. $\models (\exists x: \exists y: f) \leftrightarrow (\exists y: \exists x: f)$
7. Falls x eine Variable ist, für die $x \notin FV(f)$ ist, so haben wir

$$\models (\forall x: f) \leftrightarrow f \quad \text{und} \quad \models (\exists x: f) \leftrightarrow f.$$

8. Falls x eine Variable ist, für die $x \notin FV(g) \cup BV(g)$ gilt, so haben wir die folgenden Äquivalenzen:

- (a) $\models (\forall x: f) \vee g \leftrightarrow \forall x: (f \vee g)$
- (b) $\models g \vee (\forall x: f) \leftrightarrow \forall x: (g \vee f)$
- (c) $\models (\exists x: f) \wedge g \leftrightarrow \exists x: (f \wedge g)$
- (d) $\models g \wedge (\exists x: f) \leftrightarrow \exists x: (g \wedge f)$

Um die Äquivalenzen der letzten Gruppe anwenden zu können, ist es notwendig, gebundene Variablen umzubenennen. Ist f eine prädikatenlogische Formel und sind x und y zwei Variablen, so bezeichnet $f[x/y]$ die Formel, die aus f dadurch entsteht, dass jedes Auftreten der Variablen x in f durch y ersetzt wird. Beispielsweise gilt

$$(\forall u : \exists v : p(u, v))[u/z] = \forall z : \exists v : p(z, v)$$

Damit können wir eine letzte Äquivalenz angeben: Ist f eine prädikatenlogische Formel, ist $x \in BV(F)$ und ist y eine Variable, die in f nicht auftritt, so gilt

$$\models f \leftrightarrow f[x/y].$$

Mit Hilfe der oben stehenden Äquivalenzen können wir eine Formel so umformen, dass die Quantoren nur noch außen stehen. Eine solche Formel ist dann in *pränexer Normalform*. Wir führen das Verfahren an einem Beispiel vor: Wir zeigen, dass die Formel

$$(\forall x: p(x)) \rightarrow (\exists x: p(x))$$

allgemeingültig ist:

$$\begin{aligned} & (\forall x: p(x)) \rightarrow (\exists x: p(x)) \\ \leftrightarrow & \neg(\forall x: p(x)) \vee (\exists x: p(x)) \\ \leftrightarrow & (\exists x: \neg p(x)) \vee (\exists x: p(x)) \\ \leftrightarrow & \exists x: (\neg p(x) \vee p(x)) \\ \leftrightarrow & \exists x: \top \\ \leftrightarrow & \top \end{aligned}$$

In diesem Fall haben wir Glück gehabt, dass es uns gelungen ist, die Formel als Tautologie zu erkennen. Im Allgemeinen reichen die obigen Umformungen aber nicht aus, um prädikatenlogische Tautologien erkennen zu können. Um Formeln noch stärker normalisieren zu können, führen wir einen weiteren Äquivalenz-Begriff ein. Diesen Begriff wollen wir vorher durch ein Beispiel motivieren. Wir betrachten die beiden Formeln

$$f_1 = \forall x: \exists y: p(x, y) \quad \text{und} \quad f_2 = \forall x: p(x, s(x)).$$

Die beiden Formeln f_1 und f_2 sind nicht äquivalent, denn sie entstammen noch nicht einmal der gleichen Signatur: In der Formel f_2 wird das Funktions-Zeichen s verwendet, das in der Formel f_1 überhaupt nicht auftritt. Auch wenn die beiden Formeln f_1 und f_2 nicht äquivalent sind, so besteht zwischen ihnen doch die folgende Beziehung: Ist S_1 eine prädikatenlogische Struktur, in der die Formel f_1 gilt:

$$S_1 \models f_1,$$

dann können wir diese Struktur zu einer Struktur S_2 erweitern, in der die Formel f_2 gilt:

$$S_2 \models f_2.$$

Dazu muss lediglich die Interpretation des Funktions-Zeichens s so gewählt werden, dass für jedes x tatsächlich $p(x, s(x))$ gilt. Dies ist möglich, denn die Formel f_1 sagt ja aus, dass wir tatsächlich zu jedem x einen Wert y finden, für den $p(x, y)$ gilt. Die Funktion s muss also lediglich zu jedem x dieses y zurück geben.

Definition 54 (Skolemisierung) Es sei $\Sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, \text{arity} \rangle$ eine Signatur. Ferner sei f eine geschlossene Σ -Formel der Form

$$f = \forall x_1, \dots, x_n: \exists y: g(x_1, \dots, x_n, y).$$

Dann wählen wir ein neues n -stelliges Funktions-Zeichen s , d.h. wir nehmen ein Zeichen s , dass in der Menge \mathcal{F} nicht auftritt und erweitern die Signatur Σ zu der Signatur

$$\Sigma' := \langle \mathcal{V}, \mathcal{F} \cup \{s\}, \mathcal{P}, \text{arity} \cup \{ \langle s, n \rangle \} \rangle,$$

in der wir s als neues n -stelliges Funktions-Zeichen deklarieren. Anschließend definieren wir die Σ' -Formel f' wie folgt:

$$f' := \text{Skolem}(f) := \forall x_1: \dots \forall x_n: g(x_1, \dots, x_n, s(x_1, \dots, x_n))$$

Wir lassen also den Existenz-Quantor $\exists y$ weg und ersetzen jedes Auftreten der Variable y durch den Term $s(x_1, \dots, x_n)$. Wir sagen, dass die Formel f' aus der Formel f durch einen Skolemisierungsschritt hervorgegangen ist. \square

In welchem Sinne sind eine Formel f und eine Formel f' , die aus f durch einen Skolemisierungsschritt hervorgegangen sind, äquivalent? Zur Beantwortung dieser Frage dient die folgende Definition.

Definition 55 (Erfüllbarkeits-Äquivalenz) Zwei geschlossene Formeln f und g heißen *erfüllbarkeits-äquivalent* falls f und g entweder beide erfüllbar oder beide unerfüllbar sind. Wenn f und g erfüllbarkeits-äquivalent sind, so schreiben wir

$$f \approx_e g.$$

\square

Satz 56 Falls die Formel f' aus der Formel f durch einen Skolemisierungsschritt hervorgegangen ist, so sind f und f' erfüllbarkeits-äquivalent.

Wir können nun ein einfaches Verfahren angeben, um Existenz-Quantoren aus einer Formel zu eliminieren. Dieses Verfahren besteht aus zwei Schritten: Zunächst bringen wir die Formel in pränex Normalform. Anschließend können wir die Existenz-Quantoren der Reihe nach durch Skolemisierungsschritte eliminieren. Nach dem eben gezeigten Satz ist die resultierende Formel zu der ursprünglichen Formel erfüllbarkeits-äquivalent. Dieses Verfahren der Eliminierung von Existenz-Quantoren durch die Einführung neuer Funktions-Zeichen wird als *Skolemisierung* bezeichnet. Haben wir eine Formel F in pränex Normalform gebracht und anschließend skolemisiert, so hat das Ergebnis die Gestalt

$$\forall x_1, \dots, x_n : g$$

und in der Formel g treten keine Quantoren mehr auf. Die Formel g wird auch als die *Matrix* der obigen Formel bezeichnet. Wir können nun g mit Hilfe der uns aus dem letzten Kapitel bekannten aussagenlogischen Äquivalenzen in konjunktive Normalform bringen. Wir haben dann eine Formel der Gestalt

$$\forall x_1, \dots, x_n : (k_1 \wedge \dots \wedge k_m).$$

Dabei sind die k_i Disjunktionen von *Literals*. (In der Prädikatenlogik ist ein Literal entweder eine atomare Formel oder die Negation einer atomaren Formel.) Wenden wir hier die Äquivalenz $(\forall x: f_1 \wedge f_2) \leftrightarrow (\forall x: f_1) \wedge (\forall x: f_2)$ an, so können wir die All-Quantoren auf die einzelnen k_i verteilen und die resultierende Formel hat die Gestalt

$$(\forall x_1, \dots, x_n : k_1) \wedge \dots \wedge (\forall x_1, \dots, x_n : k_m).$$

Ist eine Formel F in der obigen Gestalt, so sagen wir, dass F in *prädikatenlogischer Klausel-Normalform* ist und eine Formel der Gestalt

$$\forall x_1, \dots, x_n : k,$$

bei der k eine Disjunktion prädikatenlogischer Literale ist, bezeichnen wir als *prädikatenlogische Klausel*. Ist M eine Menge von Formeln deren Erfüllbarkeit wir untersuchen wollen, so können wir nach dem bisher gezeigten M immer in eine Menge prädikatenlogischer Klauseln umformen. Da dann nur noch All-Quantoren vorkommen, können wir hier die Notation noch vereinfachen indem wir vereinbaren, dass alle Formeln implizit allquantifiziert sind, wir lassen also die All-Quantoren weg.

Wozu sind nun die Umformungen in Skolem-Normalform gut? Es geht darum, dass wir ein Verfahren entwickeln wollen, mit dem es möglich ist für eine prädikatenlogische Formel f zu zeigen, dass f allgemeingültig ist, dass also

$$\models f$$

gilt. Wir wissen, dass

$$\models f \quad \text{g.d.w.} \quad \{ \neg f \} \models \perp$$

gilt, denn die Formel f ist genau dann allgemeingültig, wenn es keine Struktur gibt, in der die Formel $\neg f$ erfüllbar ist. Wir bilden daher zunächst $\neg f$ und formen $\neg f$ in prädikatenlogische Klausel-Normalform um. Wir erhalten Klauseln k_1, \dots, k_n , so dass

$$\neg f \approx_e k_1 \wedge \dots \wedge k_n$$

gilt. Anschließend versuchen wir, aus den Klauseln k_1, \dots, k_n einen Widerspruch herzuleiten:

$$\{k_1, \dots, k_n\} \vdash \perp$$

Wenn dies gelingt wissen wir, dass die Menge $\{k_1, \dots, k_n\}$ unerfüllbar ist. Dann ist auch $\neg f$ unerfüllbar und damit ist f allgemeingültig. Damit wir aus den Klauseln k_1, \dots, k_n einen Widerspruch herleiten können, brauchen wir natürlich noch einen Kalkül, der mit prädikatenlogischen Klauseln arbeitet. Einen solchen Kalkül werden wir am Ende dieses Kapitels vorstellen.

Um das Verfahren näher zu erläutern demonstrieren wir es an einem Beispiel. Wir wollen untersuchen, ob

$$\models (\exists x: \forall y: p(x, y)) \rightarrow (\forall y: \exists x: p(x, y))$$

gilt. Wir wissen, dass dies äquivalent dazu ist, dass

$$\left\{ \neg \left((\exists x: \forall y: p(x, y)) \rightarrow (\forall y: \exists x: p(x, y)) \right) \right\} \models \perp$$

gilt. Wir bringen zunächst die negierte Formel in präfixe Normalform.

$$\begin{aligned} & \neg \left((\exists x: \forall y: p(x, y)) \rightarrow (\forall y: \exists x: p(x, y)) \right) \\ \leftrightarrow & \neg \left(\neg (\exists x: \forall y: p(x, y)) \vee (\forall y: \exists x: p(x, y)) \right) \\ \leftrightarrow & (\exists x: \forall y: p(x, y)) \wedge \neg (\forall y: \exists x: p(x, y)) \\ \leftrightarrow & (\exists x: \forall y: p(x, y)) \wedge (\exists y: \neg \exists x: p(x, y)) \\ \leftrightarrow & (\exists x: \forall y: p(x, y)) \wedge (\exists y: \forall x: \neg p(x, y)) \end{aligned}$$

Um an dieser Stelle weitermachen zu können, ist es nötig, die Variablen in dem zweiten Glied der Konjunktion umzubenennen. Wir ersetzen x durch u und y durch v und erhalten

$$\begin{aligned} & (\exists x: \forall y: p(x, y)) \wedge (\exists y: \forall x: \neg p(x, y)) \\ \leftrightarrow & (\exists x: \forall y: p(x, y)) \wedge (\exists v: \forall u: \neg p(u, v)) \\ \leftrightarrow & \exists v: \left((\exists x: \forall y: p(x, y)) \wedge (\forall u: \neg p(u, v)) \right) \\ \leftrightarrow & \exists v: \exists x: \left((\forall y: p(x, y)) \wedge (\forall u: \neg p(u, v)) \right) \\ \leftrightarrow & \exists v: \exists x: \forall y: \left(p(x, y) \wedge (\forall u: \neg p(u, v)) \right) \\ \leftrightarrow & \exists v: \exists x: \forall y: \forall u: \left(p(x, y) \wedge \neg p(u, v) \right) \end{aligned}$$

An dieser Stelle müssen wir skolemisieren um die Existenz-Quantoren los zu werden. Wir führen dazu zwei neue Funktions-Zeichen s_1 und s_2 ein. Dabei gilt $\text{arity}(s_1) = 0$ und $\text{arity}(s_2) = 0$, denn vor den Existenz-Quantoren stehen keine All-Quantoren.

$$\begin{aligned} & \exists v: \exists x: \forall y: \forall u: \left(p(x, y) \wedge \neg p(u, v) \right) \\ \approx_e & \exists x: \forall y: \forall u: \left(p(x, y) \wedge \neg p(u, s_1) \right) \\ \approx_e & \forall y: \forall u: \left(p(s_2, y) \wedge \neg p(u, s_1) \right) \end{aligned}$$

Da jetzt nur noch All-Quantoren auftreten, können wir diese auch noch weglassen, da wir ja vereinbart haben, dass alle freien Variablen implizit allquantifiziert sind. Damit können wir nun die prädikatenlogische Klausel-Normalform angeben, diese ist

$$M := \left\{ \{p(s_2, y)\}, \{\neg p(u, s_1)\} \right\}.$$

Wir zeigen, dass die Menge M widersprüchlich ist. Dazu betrachten wir zunächst die Klausel $\{p(s_2, y)\}$ und setzen in dieser Klausel für y die Konstante s_1 ein. Damit erhalten wir die Klausel

$$\{p(s_2, s_1)\}. \tag{1}$$

Das Ersetzen von y durch s_1 begründen wir damit, dass die obige Klausel ja implizit allquantifiziert ist und wenn etwas für alle y gilt, dann sicher auch für $y = s_1$.

Als nächstes betrachten wir die Klausel $\{\neg p(u, s_1)\}$. Hier setzen wir für die Variablen u die Konstante s_2 ein und erhalten dann die Klausel

$$\{\neg p(s_2, s_1)\} \quad (2)$$

Nun wenden wir auf die Klauseln (1) und (2) die Schnitt-Regel an und finden

$$\{p(s_2, s_1)\}, \quad \{\neg p(s_2, s_1)\} \quad \vdash \quad \{\}.$$

Damit haben wir einen Widerspruch hergeleitet und gezeigt, dass die Menge M unerfüllbar ist. Damit ist dann auch

$$\left\{ \neg \left((\exists x: \forall y: p(x, y)) \rightarrow (\forall y: \exists x: p(x, y)) \right) \right\}$$

unerfüllbar und folglich gilt

$$\models (\exists x: \forall y: p(x, y)) \rightarrow (\forall y: \exists x: p(x, y)).$$

6.4 Unifikation

In dem Beispiel im letzten Abschnitt haben wir die Terme s_1 und s_2 geraten, die wir für die Variablen y und u in den Klauseln $\{p(s_2, y)\}$ und $\{\neg p(u, s_1)\}$ eingesetzt haben. Wir haben diese Terme mit dem Ziel gewählt, später die Schnitt-Regel anwenden zu können. In diesem Abschnitt zeigen wir nun ein Verfahren, mit dessen Hilfe wir die benötigten Terme ausrechnen können. Dazu benötigen wir zunächst den Begriff einer *Substitution*.

Definition 57 (Substitution) Es sei eine Signatur

$$\Sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, \text{arity} \rangle$$

gegeben. Eine Σ -Substitution ist eine endliche Menge von Paaren der Form

$$\sigma = \{ \langle x_1, t_1 \rangle, \dots, \langle x_n, t_n \rangle \}.$$

Dabei gilt:

1. $x_i \in \mathcal{V}$, die x_i sind also Variablen.
2. $t_i \in \mathcal{T}_\Sigma$, die t_i sind also Terme.
3. Für $i \neq j$ ist $x_i \neq x_j$, die Variablen sind also paarweise verschieden.

Ist $\sigma = \{ \langle x_1, t_1 \rangle, \dots, \langle x_n, t_n \rangle \}$ eine Σ -Substitution, so schreiben wir

$$\sigma = [x_1 \mapsto t_1, \dots, x_n \mapsto t_n].$$

Außerdem definieren wir den *Domain* einer Substitution als

$$\text{dom}(\sigma) = \{x_1, \dots, x_n\}.$$

Die Menge aller Substitutionen bezeichnen wir mit *Subst*. □

Substitutionen werden für uns dadurch interessant, dass wir sie auf Terme *anwenden* können. Ist t ein Term und σ eine Substitution, so ist $t\sigma$ der Term, der aus t dadurch entsteht, dass jedes Vorkommen einer Variablen x_i durch den zugehörigen Term t_i ersetzt wird. Die formale Definition folgt.

Definition 58 (Anwendung einer Substitution)

Es sei t ein Term und es sei $\sigma = [x_1 \mapsto t_1, \dots, x_n \mapsto t_n]$ eine Substitution. Wir definieren die *Anwendung* von σ auf t (Schreibweise $t\sigma$) durch Induktion über den Aufbau von t :

1. Falls t eine Variable ist, gibt es zwei Fälle:

- (a) $t = x_i$ für ein $i \in \{1, \dots, n\}$. Dann definieren wir $x_i\sigma := t_i$.

(b) $t = y$ mit $y \in \mathcal{V}$, aber $y \notin \{x_1, \dots, x_n\}$. Dann definieren wir $y\sigma := y$.

2. Andernfalls muß t die Form $t = f(s_1, \dots, s_m)$ haben. Dann können wir $t\sigma$ durch

$$f(s_1, \dots, s_m)\sigma := f(s_1\sigma, \dots, s_m\sigma).$$

definieren, denn nach Induktions-Voraussetzung sind die Ausdrücke $s_i\sigma$ bereits definiert. \square

Genau wie wir Substitutionen auf Terme anwenden können, können wir eine Substitution auch auf prädikatenlogische Klauseln anwenden. Dabei werden Prädikats-Zeichen und Junktoren wie Funktions-Zeichen behandelt. Wir ersparen uns eine formale Definition und geben statt dessen zunächst einige Beispiele. Wir definieren eine Substitution σ durch

$$\sigma := [x_1 \mapsto c, x_2 \mapsto f(d)].$$

In den folgenden drei Beispielen demonstrieren wir zunächst, wie eine Substitution auf einen Term angewendet werden kann. Im vierten Beispiel wenden wir die Substitution dann auf eine Formel an:

1. $x_3\sigma = x_3$,
2. $f(x_2)\sigma = f(f(d))$,
3. $h(x_1, g(x_2))\sigma = h(c, g(f(d)))$.
4. $\{p(x_2), q(d, h(x_3, x_1))\}\sigma = \{p(f(d)), q(d, h(x_3, c))\}$.

Als nächstes zeigen wir, wie Substitutionen miteinander verknüpft werden können.

Definition 59 (Komposition von Substitutionen) Es seien

$$\sigma = [x_1 \mapsto s_1, \dots, x_m \mapsto s_m] \quad \text{und} \quad \tau = [y_1 \mapsto t_1, \dots, y_n \mapsto t_n]$$

zwei Substitutionen mit $\text{dom}(\sigma) \cap \text{dom}(\tau) = \{\}$. Dann definieren wir die *Komposition* $\sigma\tau$ von σ und τ als

$$\sigma\tau := [x_1 \mapsto s_1\tau, \dots, x_m \mapsto s_m\tau, y_1 \mapsto t_1, \dots, y_n \mapsto t_n] \quad \square$$

Beispiel: Wir führen das obige Beispiel fort und setzen

$$\sigma := [x_1 \mapsto c, x_2 \mapsto f(x_3)] \quad \text{und} \quad \tau := [x_3 \mapsto h(c, c), x_4 \mapsto d].$$

Dann gilt:

$$\sigma\tau = [x_1 \mapsto c, x_2 \mapsto f(h(c, c)), x_3 \mapsto h(c, c), x_4 \mapsto d]. \quad \square$$

Die Definition der Komposition von Substitutionen ist mit dem Ziel gewählt worden, dass der folgende Satz gilt.

Satz 60 Ist t ein Term und sind σ und τ Substitutionen mit $\text{dom}(\sigma) \cap \text{dom}(\tau) = \{\}$, so gilt

$$(t\sigma)\tau = t(\sigma\tau). \quad \square$$

Der Satz kann durch Induktion über den Aufbau des Termes t bewiesen werden.

Definition 61 (Syntaktische Gleichung) Unter einer *syntaktischen Gleichung* verstehen wir in diesem Abschnitt ein Konstrukt der Form $s \doteq t$, wobei einer der beiden folgenden Fälle vorliegen muß:

1. s und t sind Terme oder
2. s und t sind atomare Formeln.

Weiter definieren wir ein *syntaktisches Gleichungs-System* als eine Menge von syntaktischen Gleichungen. \square

Was syntaktische Gleichungen angeht machen wir keinen Unterschied zwischen Funktions-Zeichen und Prädikats-Zeichen. Dieser Ansatz ist deswegen berechtigt, weil wir Prädikats-Zeichen ja auch als spezielle Funktions-Zeichen auffassen können, nämlich als Funktions-Zeichen, die einen Wahrheitswert aus der Menge \mathbb{B} berechnen.

Definition 62 (Unifikator) Eine Substitution σ *löst* eine syntaktische Gleichung $s \doteq t$ genau dann, wenn $s\sigma = t\sigma$ ist, wenn also durch die Anwendung von σ auf s und t tatsächlich identische Objekte entstehen. Ist E ein syntaktisches Gleichungs-System, so sagen wir, dass σ ein *Unifikator* von E ist wenn σ jede syntaktische Gleichung in E löst. \square

Ist $E = \{s_1 \doteq t_1, \dots, s_n \doteq t_n\}$ eine syntaktisches Gleichungs-System und ist σ eine Substitution, so definieren wir

$$E\sigma := \{s_1\sigma \doteq t_1\sigma, \dots, s_n\sigma \doteq t_n\sigma\}.$$

Beispiel: Wir verdeutlichen die bisher eingeführten Begriffe anhand eines Beispiels. Wir betrachten die Gleichung

$$p(x_1, f(x_4)) \doteq p(x_2, x_3)$$

und definieren die Substitution

$$\sigma := [x_1 \mapsto x_2, x_3 \mapsto f(x_4)].$$

Die Substitution σ löst die obige syntaktische Gleichung, denn es gilt

$$\begin{aligned} p(x_1, f(x_4))\sigma &= p(x_2, f(x_4)) \quad \text{und} \\ p(x_2, x_3)\sigma &= p(x_2, f(x_4)). \end{aligned}$$

Als nächstes entwickeln wir ein Verfahren, mit dessen Hilfe wir von einer vorgegebenen Menge E von syntaktischen Gleichungen entscheiden können, ob es einen Unifikator σ für E gibt. Wir überlegen uns zunächst, in welchen Fällen wir eine syntaktischen Gleichung $s \doteq t$ garantiert nicht lösen können. Da gibt es zwei Möglichkeiten: Eine syntaktische Gleichung

$$f(s_1, \dots, s_m) \doteq g(t_1, \dots, t_n)$$

ist sicher dann nicht durch eine Substitution lösbar, wenn f und g verschiedene Funktions-Zeichen sind, denn für jede Substitution σ gilt ja

$$f(s_1, \dots, s_m)\sigma = f(s_1\sigma, \dots, s_m\sigma) \quad \text{und} \quad g(t_1, \dots, t_n)\sigma = g(t_1\sigma, \dots, t_n\sigma).$$

Falls $f \neq g$ ist, haben die Terme $f(s_1, \dots, s_m)\sigma$ und $g(t_1, \dots, t_n)\sigma$ verschieden Funktions-Zeichen und können daher syntaktisch nicht identisch werden.

Die andere Form einer syntaktischen Gleichung, die garantiert unlösbar ist, ist

$$x \doteq f(t_1, \dots, t_n) \quad \text{falls } x \in \text{Var}(f(t_1, \dots, t_n)).$$

Das diese syntaktische Gleichung unlösbar ist liegt daran, dass die rechte Seite immer mindestens ein Funktions-Zeichen mehr enthält als die linke.

Mit diesen Vorbemerkungen können wir nun ein Verfahren angeben, mit dessen Hilfe es möglich ist, Mengen von syntaktischen Gleichungen zu lösen, oder festzustellen, dass es keine Lösung gibt. Das Verfahren operiert auf Paaren der Form $\langle F, \tau \rangle$. Dabei ist F ein syntaktisches Gleichungs-System und τ ist eine Substitution. Wir starten das Verfahren mit dem Paar $\langle E, [] \rangle$. Hierbei ist E das zu lösende Gleichungs-System und $[]$ ist die leere Substitution. Das Verfahren arbeitet indem die im folgenden dargestellten Reduktions-Regeln solange angewendet werden, bis entweder feststeht, dass die Menge der Gleichungen keine Lösung hat, oder aber ein Paar der Form $\langle \{\}, \sigma \rangle$ erreicht wird. In diesem Fall ist σ ein Unifikator der Menge E , mit der wir gestartet sind. Es folgen die Reduktions-Regeln:

1. Falls $y \in \mathcal{V}$ eine Variable ist, die nicht in dem Term t auftritt, so können wir die folgende Reduktion durchführen:

$$\langle E \cup \{y \doteq t\}, \sigma \rangle \rightsquigarrow \langle E[y \mapsto t], \sigma[y \mapsto t] \rangle$$

Diese Reduktions-Regel ist folgendermaßen zu lesen: Enthält die zu untersuchende Menge von syntaktischen Gleichungen eine Gleichung der Form $y \doteq t$, wobei die Variable y nicht in t auftritt, dann können wir

diese Gleichung aus der gegebenen Menge von Gleichungen entfernen. Gleichzeitig wird die Substitution σ in die Substitution $\sigma[y \mapsto t]$ transformiert und auf die restlichen syntaktischen Gleichungen wird die Substitution $[y \mapsto t]$ angewendet.

2. Wenn die Variable y in dem Term t auftritt, falls also $y \in \text{var}(t)$ ist und wenn außerdem $t \neq y$ ist, dann hat das Gleichungs-System $E \cup \{y \doteq t\}$ keine Lösung, wir schreiben

$$\langle E \cup \{y \doteq t\}, \sigma \rangle \rightsquigarrow \Omega.$$

3. Falls $y \in \mathcal{V}$ eine Variable ist und t keine Variable ist, so haben wir folgende Reduktions-Regel:

$$\langle E \cup \{t \doteq y\}, \sigma \rangle \rightsquigarrow \langle E \cup \{y \doteq t\}, \sigma \rangle.$$

Diese Regel wird benötigt, um anschließend eine der ersten beiden Regeln anwenden zu können.

4. Triviale syntaktische Gleichungen von Variablen können wir einfach weglassen:

$$\langle E \cup \{x \doteq x\}, \sigma \rangle \rightsquigarrow \langle E, \sigma \rangle.$$

5. Ist f ein n -stelliges Funktions-Zeichen, so gilt

$$\langle E \cup \{f(s_1, \dots, s_n) \doteq f(t_1, \dots, t_n)\}, \sigma \rangle \rightsquigarrow \langle E \cup \{s_1 \doteq t_1, \dots, s_n \doteq t_n\}, \sigma \rangle.$$

Eine syntaktische Gleichung der Form $f(s_1, \dots, s_n) \doteq f(t_1, \dots, t_n)$ wird also ersetzt durch die n syntaktischen Gleichungen $s_1 \doteq t_1, \dots, s_n \doteq t_n$.

Diese Regel ist im Übrigen der Grund dafür, dass wir mit Mengen von syntaktischen Gleichungen arbeiten müssen, denn auch wenn wir mit nur einer syntaktischen Gleichung starten, kann durch die Anwendung dieser Regel die Zahl der syntaktischen Gleichungen erhöht werden.

Ein Spezialfall dieser Regel ist

$$\langle E \cup \{c \doteq c\}, \sigma \rangle \rightsquigarrow \langle E, \sigma \rangle.$$

Hier steht c für eine Konstante, also ein 0-stelliges Funktions-Zeichen. Triviale Gleichungen über Konstanten können also einfach weggelassen werden.

6. Das Gleichungs-System $E \cup \{f(s_1, \dots, s_m) \doteq g(t_1, \dots, t_n)\}$ hat keine Lösung, falls die Funktions-Zeichen f und g verschieden sind, wir schreiben

$$\langle E \cup \{f(s_1, \dots, s_m) \doteq g(t_1, \dots, t_n)\}, \sigma \rangle \rightsquigarrow \Omega \quad \text{falls } f \neq g.$$

Haben wir ein nicht-leeres Gleichungs-System E gegeben und starten mit dem Paar $\langle E, [] \rangle$, so läßt sich immer eine der obigen Regeln anwenden. Diese geht solange bis einer der folgenden Fälle eintritt:

1. Die 2. oder 6. Regel ist anwendbar. Dann ist das Ergebnis Ω und das Gleichungs-System E hat keine Lösung.
2. Das Paar $\langle E, [] \rangle$ wird reduziert zu einem Paar $\langle \{\}, \sigma \rangle$. Dann ist σ ein Unifikator von E . In diesem Falls schreiben wir $\sigma = \text{mgu}(E)$. Falls $E = \{s \doteq t\}$ ist, schreiben wir auch $\sigma = \text{mgu}(s, t)$. Die Abkürzung mgu steht hier für “*most general unifier*”.

Beispiel: Wir wenden das oben dargestellte Verfahren an, um die syntaktische Gleichung

$$p(x_1, f(x_4)) \doteq p(x_2, x_3)$$

zu lösen. Wir haben die folgenden Reduktions-Schritte:

$$\begin{aligned} & \langle \{p(x_1, f(x_4)) \doteq p(x_2, x_3)\}, [] \rangle \\ \rightsquigarrow & \langle \{x_1 \doteq x_2, f(x_4) \doteq x_3\}, [] \rangle \\ \rightsquigarrow & \langle \{f(x_4) \doteq x_3\}, [x_1 \mapsto x_2] \rangle \\ \rightsquigarrow & \langle \{x_3 \doteq f(x_4)\}, [x_1 \mapsto x_2] \rangle \\ \rightsquigarrow & \langle \{\}, [x_1 \mapsto x_2, x_3 \mapsto f(x_4)] \rangle \end{aligned}$$

In diesem Fall ist das Verfahren also erfolgreich und wir erhalten die Substitution

$$[x_1 \mapsto x_2, x_3 \mapsto f(x_4)]$$

als Lösung der oben gegebenen syntaktischen Gleichung.

Wir geben ein weiteres Beispiel und betrachten das Gleichungs-System

$$E = \{p(h(x_1, c)) \doteq p(x_2), q(x_2, d) \doteq q(h(d, c), x_4)\}$$

Wir haben folgende Reduktions-Schritte:

$$\begin{aligned} & \langle \{p(h(x_1, c)) \doteq p(x_2), q(x_2, d) \doteq q(h(d, c), x_4)\}, [] \rangle \\ \rightsquigarrow & \langle \{p(h(x_1, c)) \doteq p(x_2), x_2 \doteq h(d, c), d \doteq x_4\}, [] \rangle \\ \rightsquigarrow & \langle \{p(h(x_1, c)) \doteq p(x_2), x_2 \doteq h(d, c), x_4 \doteq d\}, [] \rangle \\ \rightsquigarrow & \langle \{p(h(x_1, c)) \doteq p(x_2), x_2 \doteq h(d, c)\}, [x_4 \mapsto d] \rangle \\ \rightsquigarrow & \langle \{p(h(x_1, c)) \doteq p(h(d, c))\}, [x_4 \mapsto d, x_2 \mapsto h(d, c)] \rangle \\ \rightsquigarrow & \langle \{h(x_1, c) \doteq h(d, c)\}, [x_4 \mapsto d, x_2 \mapsto h(d, c)] \rangle \\ \rightsquigarrow & \langle \{x_1 \doteq d, c \doteq c\}, [x_4 \mapsto d, x_2 \mapsto h(d, c)] \rangle \\ \rightsquigarrow & \langle \{x_1 \doteq d\}, [x_4 \mapsto d, x_2 \mapsto h(d, c)] \rangle \\ \rightsquigarrow & \langle \{\}, [x_4 \mapsto d, x_2 \mapsto h(d, c), x_1 \mapsto d] \rangle \end{aligned}$$

Damit haben wir die Substitution $[x_4 \mapsto d, x_2 \mapsto h(d, c), x_1 \mapsto d]$ als Lösung des anfangs gegebenen syntaktischen Gleichungs-Systems gefunden. \square

6.5 Ein Kalkül für die Prädikatenlogik

Der Kalkül, den wir in diesem Abschnitt für die Prädikatenlogik einführen, besteht aus zwei Schluß-Regeln, die wir jetzt definieren.

Definition 63 (Resolution) Es gelte:

1. k_1 und k_2 sind prädikatenlogische Klauseln,
2. $p(s_1, \dots, s_n)$ und $p(t_1, \dots, t_n)$ sind atomare Formeln,
3. die syntaktische Gleichung $p(s_1, \dots, s_n) \doteq p(t_1, \dots, t_n)$ ist lösbar mit

$$\mu = \text{mgu}(p(s_1, \dots, s_n), p(t_1, \dots, t_n)).$$

Dann ist

$$\frac{k_1 \cup \{p(s_1, \dots, s_n)\} \quad \{\neg p(t_1, \dots, t_n)\} \cup k_2}{k_1\mu \cup k_2\mu}$$

eine Anwendung der *Resolutions-Regel*. □

Die Resolutions-Regel ist eine Kombination aus der *Substitutions-Regel* und der Schnitt-Regel. Die Substitutions-Regel hat die Form

$$\frac{k}{k\sigma}.$$

Hierbei ist k eine prädikatenlogische Klausel und σ ist eine Substitution. Unter Umständen kann es sein, dass wir bei der Anwendung der Resolutions-Regel die Variablen in einer der beiden Klauseln erst umbenennen müssen bevor wir die Regel anwenden können. Betrachten wir dazu ein Beispiel. Die Klausel-Menge

$$M = \left\{ \{p(x)\}, \{\neg p(f(x))\} \right\}$$

ist widersprüchlich. Wir können die Resolutions-Regel aber nicht unmittelbar anwenden, denn die syntaktische Gleichung

$$p(x) \doteq p(f(x))$$

ist unlösbar. Das liegt daran, dass **zufällig** in beiden Klauseln die selbe Variable verwendet wird. Wenn wir die Variable x in der zweiten Klausel jedoch zu y umbenennen, erhalten wir die Klausel-Menge

$$\left\{ \{p(x)\}, \{\neg p(f(y))\} \right\}.$$

Hier können wir die Resolutions-Regel anwenden, denn die syntaktische Gleichung

$$p(x) \doteq p(f(y))$$

hat die Lösung $[x \mapsto f(y)]$. Dann erhalten wir

$$\{p(x)\}, \quad \{\neg p(f(y))\} \quad \vdash \quad \{\}.$$

und haben damit die Inkonsistenz der Klausel-Menge M nachgewiesen.

Die Resolutions-Regel alleine ist nicht ausreichend, um aus einer Klausel-Menge M , die inkonsistent ist, in jedem Fall die leere Klausel ableiten zu können: Wir brauchen noch eine zweite Regel. Um das einzusehen, betrachten wir die Klausel-Menge

$$M = \left\{ \{p(f(x), y), p(u, g(v))\}, \{\neg p(f(x), y), \neg p(u, g(v))\} \right\}$$

Wir werden gleich zeigen, dass die Menge M widersprüchlich ist. Man kann nachweisen, dass mit der Resolutions-Regel alleine ein solcher Nachweis nicht gelingt. Ein einfacher, aber für die Vorlesung zu aufwendiger Nachweis dieser Behauptung kann geführt werden, indem wir ausgehend von der Menge M alle möglichen Resolutions-Schritte durchführen. Dabei würden wir dann sehen, dass die leere Klausel nie berechnet wird. Wir stellen daher jetzt die Faktorisierungs-Regel vor, mir der wir später zeigen werden, dass M widersprüchlich ist.

Definition 64 (Faktorisierung) *Es gelte*

1. k ist eine prädikatenlogische Klausel,
2. $p(s_1, \dots, s_n)$ und $p(t_1, \dots, t_n)$ sind atomare Formeln,
3. die syntaktische Gleichung $p(s_1, \dots, s_n) \doteq p(t_1, \dots, t_n)$ ist lösbar,
4. $\mu = mgu(p(s_1, \dots, s_n), p(t_1, \dots, t_n))$.

Dann sind

$$\frac{k \cup \{p(s_1, \dots, s_n), p(t_1, \dots, t_n)\}}{k\mu \cup \{p(s_1, \dots, s_n)\mu\}} \quad \text{und} \quad \frac{k \cup \{\neg p(s_1, \dots, s_n), \neg p(t_1, \dots, t_n)\}}{k\mu \cup \{\neg p(s_1, \dots, s_n)\mu\}}$$

Anwendungen der *Faktorisierungs-Regel*. □

Wir zeigen, wie sich mit Resolutions- und Faktorisierungs-Regel die Widersprüchlichkeit der Menge M beweisen läßt.

1. Zunächst wenden wir die Faktorisierungs-Regel auf die erste Klausel an. Dazu berechnen wir den Unifikator

$$\mu = \text{mgu}(p(f(x), y), p(u, g(v))) = [y \mapsto g(v), u \mapsto f(x)].$$

Damit können wir die Faktorisierungs-Regel anwenden:

$$\{p(f(x), y), p(u, g(v))\} \quad \vdash \quad \{p(f(x), g(v))\}.$$

2. Jetzt wenden wir die Faktorisierungs-Regel auf die zweite Klausel an. Dazu berechnen wir den Unifikator

$$\mu = \text{mgu}(\neg p(f(x), y), \neg p(u, g(v))) = [y \mapsto g(v), u \mapsto f(x)].$$

Damit können wir die Faktorisierungs-Regel anwenden:

$$\{\neg p(f(x), y), \neg p(u, g(v))\} \quad \vdash \quad \{\neg p(f(x), g(v))\}.$$

3. Wir schließen den Beweis mit einer Anwendung der Resolutions-Regel ab. Der dabei verwendete Unifikator ist die leere Substitution, es gilt also $\mu = []$.

$$\{p(f(x), g(v))\}, \quad \{\neg p(f(x), g(v))\} \quad \vdash \quad \{\}.$$

Ist M eine Menge von prädikatenlogischen Klauseln und ist k eine prädikatenlogische Klausel, die durch Anwendung der Resolutions-Regel und der Faktorisierungs-Regel aus M hergeleitet werden kann, so schreiben wir

$$M \vdash k.$$

Dies wird als M *leitet k her* gelesen.

Definition 65 (Allabschluß) Ist k eine prädikatenlogische Klausel und ist $\{x_1, \dots, x_n\}$ die Menge aller Variablen, die in k auftreten, so definieren wir den *Allabschluß* $\forall(k)$ der Klausel k als

$$\forall(k) := \forall x_1, \dots, x_n: k.$$

Die für uns wesentlichen Eigenschaften des Beweis-Begriffs $M \vdash k$ werden in den folgenden beiden Sätzen zusammengefaßt.

Satz 66 (Korrektheits-Satz)

Ist $M = \{k_1, \dots, k_n\}$ eine Menge von Klauseln und gilt $M \vdash k$, so folgt

$$\models \forall(k_1) \wedge \dots \wedge \forall(k_n) \rightarrow \forall(k).$$

Falls also eine Klausel k aus einer Menge M hergeleitet werden kann, so ist k tatsächlich eine Folgerung aus M . □

Die Umkehrung des obigen Korrektheits-Satzes gilt nur für die leere Klausel.

Satz 67 (Widerlegungs-Vollständigkeit)

Ist $M = \{k_1, \dots, k_n\}$ eine Menge von Klauseln und gilt $\models \forall(k_1) \wedge \dots \wedge \forall(k_n) \rightarrow \perp$, so folgt

$$M \vdash \{\}.$$
□

Damit haben wir nun ein Verfahren in der Hand, um für eine gegebene prädikatenlogischer Formel f die Frage, ob $\models f$ gilt, untersuchen zu können.

1. Wir berechnen zunächst die Skolem-Normalform von $\neg f$ und erhalten dabei so etwas wie

$$\neg f \approx_e \forall x_1, \dots, x_m. g.$$

2. Anschließend bringen wir die Matrix g in konjunktive Normalform:

$$g \leftrightarrow k_1 \wedge \dots \wedge k_n.$$

Daher haben wir nun

$$\neg f \approx_e k_1 \wedge \dots \wedge k_n$$

und es gilt:

$$\models f \quad \text{g.d.w.} \quad \{\neg f\} \models \perp \quad \text{g.d.w.} \quad \{k_1, \dots, k_n\} \models \perp.$$

3. Nach dem Korrektheits-Satz und dem Satz über die Widerlegungs-Vollständigkeit gilt

$$\{k_1, \dots, k_n\} \models \perp \quad \text{g.d.w.} \quad \{k_1, \dots, k_n\} \vdash \perp.$$

Wir versuchen also, nun die Widersprüchlichkeit der Menge $M = \{k_1, \dots, k_n\}$ zu zeigen, indem wir aus M die leere Klausel ableiten. Wenn diese gelingt, haben wir damit die Allgemeingültigkeit der ursprünglich gegebenen Formel f gezeigt.

Zum Abschluß demonstrieren wir das skizzierte Verfahren an einem Beispiel. Wir gehen von folgenden Axiomen aus:

1. Jeder Drache ist glücklich, wenn alle seine Kinder fliegen können.
2. Rote Drachen können fliegen.
3. Die Kinder eines roten Drachens sind immer rot.

Wie werden zeigen, dass aus diesen Axiomen folgt, dass alle roten Drachen glücklich sind. Als erstes formalisieren wir die Axiome und die Behauptung in der Prädikatenlogik. Wir wählen die folgende Signatur

$$\Sigma_{\text{Drache}} := \langle \mathcal{V}, \mathcal{F}, \mathcal{P}, \text{arity} \rangle \quad \text{mit}$$

1. $\mathcal{V} := \{x, y, z\}$.
2. $\mathcal{F} = \{\}$.
3. $\mathcal{P} := \{\text{rot}, \text{fliegt}, \text{glücklich}, \text{kind}\}$.
4. $\text{arity} := \{\langle \text{rot}, 1 \rangle, \langle \text{fliegt}, 1 \rangle, \langle \text{glücklich}, 1 \rangle, \langle \text{kind}, 2 \rangle\}$

Das Prädikat $\text{kind}(x, y)$ soll genau dann wahr sein, wenn x ein Kind von y ist. Formalisieren wir die Axiome und die Behauptung, so erhalten wir die folgenden Formeln f_1, \dots, f_4 :

1. $f_1 := \forall x : (\forall y : (\text{kind}(y, x) \rightarrow \text{fliegt}(y)) \rightarrow \text{glücklich}(x))$
2. $f_2 := \forall x : (\text{rot}(x) \rightarrow \text{fliegt}(x))$
3. $f_3 := \forall x : (\text{rot}(x) \rightarrow \forall y : (\text{kind}(y, x) \rightarrow \text{rot}(y)))$
4. $f_4 := \forall x : (\text{rot}(x) \rightarrow \text{glücklich}(x))$

Wir wollen zeigen, dass die Formel

$$f := f_1 \wedge f_2 \wedge f_3 \rightarrow f_4$$

allgemeingültig ist. Wir betrachten also die Formel $\neg f$ und stellen fest

$$\neg f \leftrightarrow f_1 \wedge f_2 \wedge f_3 \wedge \neg f_4.$$

Als nächstes müssen wir diese Formel in eine Menge von Klauseln umformen. Da es sich hier um eine Konjunktion mehrerer Formeln handelt, können wir die einzelnen Formeln f_1 , f_2 , f_3 und $\neg f_4$ getrennt in Klauseln umwandeln.

1. Die Formel f_1 kann wie folgt umgeformt werden:

$$\begin{aligned} f_1 &= \forall x : \left(\forall y : (kind(y, x) \rightarrow fliegt(y)) \rightarrow glücklich(x) \right) \\ &\leftrightarrow \forall x : \left(\neg \forall y : (kind(y, x) \rightarrow fliegt(y)) \vee glücklich(x) \right) \\ &\leftrightarrow \forall x : \left(\neg \forall y : (\neg kind(y, x) \vee fliegt(y)) \vee glücklich(x) \right) \\ &\leftrightarrow \forall x : \left(\exists y : \neg (\neg kind(y, x) \vee fliegt(y)) \vee glücklich(x) \right) \\ &\leftrightarrow \forall x : \left(\exists y : (kind(y, x) \wedge \neg fliegt(y)) \vee glücklich(x) \right) \\ &\leftrightarrow \forall x : \exists y : \left((kind(y, x) \wedge \neg fliegt(y)) \vee glücklich(x) \right) \\ &\approx_e \forall x : \left((kind(s(x), x) \wedge \neg fliegt(s(x))) \vee glücklich(x) \right) \end{aligned}$$

Im letzten Schritt haben wir dabei die Skolem-Funktion s mit $arity(s) = 1$ eingeführt. Anschaulich berechnet diese Funktion für jeden Drachen x , der nicht glücklich ist, ein Kind $s(x)$, das nicht fliegen kann. Wenn wir in der Matrix dieser Formel das “ \vee ” noch ausmultiplizieren, so erhalten wir die beiden Klauseln

$$\begin{aligned} k_1 &:= \{ kind(s(x), x), glücklich(x) \}, \\ k_2 &:= \{ \neg fliegt(s(x)), glücklich(x) \}. \end{aligned}$$

2. Analog finden wir für f_2 :

$$\begin{aligned} f_2 &= \forall x : (rot(x) \rightarrow fliegt(x)) \\ &\leftrightarrow \forall x : (\neg rot(x) \vee fliegt(x)) \end{aligned}$$

Damit ist f_2 zu folgender Klauseln äquivalent:

$$k_3 := \{ \neg rot(x), fliegt(x) \}.$$

3. Für f_3 sehen wir:

$$\begin{aligned} f_3 &= \forall x : \left(rot(x) \rightarrow \forall y : (kind(y, x) \rightarrow rot(y)) \right) \\ &\leftrightarrow \forall x : \left(\neg rot(x) \vee \forall y : (\neg kind(y, x) \vee rot(y)) \right) \\ &\leftrightarrow \forall x : \forall y : (\neg rot(x) \vee \neg kind(y, x) \vee rot(y)) \end{aligned}$$

Das liefert die folgende Klausel:

$$k_4 := \{ \neg rot(x), \neg kind(y, x), rot(y) \}.$$

4. Umformung der Negation von f_4 liefert:

$$\begin{aligned} \neg f_4 &= \neg \forall x : (rot(x) \rightarrow glücklich(x)) \\ &\leftrightarrow \neg \forall x : (\neg rot(x) \vee glücklich(x)) \\ &\leftrightarrow \exists x : \neg (\neg rot(x) \vee glücklich(x)) \\ &\leftrightarrow \exists x : (rot(x) \wedge \neg glücklich(x)) \\ &\approx_e rot(d) \wedge \neg glücklich(d) \end{aligned}$$

Die hier eingeführte Skolem-Konstante d steht für einen unglücklichen roten Drachen. Das führt zu den Klauseln

$$k_5 = \{rot(d)\},$$

$$k_6 = \{\neg glücklich(d)\}.$$

Wir müssen also untersuchen, ob die Menge M , die aus den folgenden Klauseln besteht, widersprüchlich ist:

1. $k_1 = \{kind(s(x), x), glücklich(x)\}$
2. $k_2 = \{\neg fliegt(s(x)), glücklich(x)\}$
3. $k_3 = \{\neg rot(x), fliegt(x)\}$
4. $k_4 = \{\neg rot(x), \neg kind(y, x), rot(y)\}$
5. $k_5 = \{rot(d)\}$
6. $k_6 = \{\neg glücklich(d)\}$

Sei also $M := \{k_1, k_2, k_3, k_4, k_5, k_6\}$. Wir zeigen, dass $M \vdash \perp$ gilt:

1. Es gilt

$$mgu(rot(d), rot(x)) = [x \mapsto d].$$

Daher können wir die Resolutions-Regel auf die Klauseln k_5 und k_4 wie folgt anwenden:

$$\{rot(d)\}, \{\neg rot(x), \neg kind(y, x), rot(y)\} \vdash \{\neg kind(y, d), rot(y)\}.$$

2. Wir wenden nun auf die resultierende Klausel und auf die Klausel k_1 die Resolutions-Regel an. Dazu berechnen wir zunächst

$$mgu(kind(y, d), kind(s(x), x)) = [y \mapsto s(d), x \mapsto d].$$

Dann haben wir

$$\{\neg kind(y, d), rot(y)\}, \{kind(s(x), x), glücklich(x)\} \vdash \{glücklich(d), rot(s(d))\}.$$

3. Jetzt wenden wir auf die eben abgeleitete Klausel und die Klausel k_6 die Resolutions-Regel an. Wir haben:

$$mgu(glücklich(d), glücklich(d)) = []$$

Also erhalten wir

$$\{glücklich(d), rot(s(d))\}, \{\neg glücklich(d)\} \vdash \{rot(s(d))\}.$$

4. Auf die Klausel $\{rot(s(d))\}$ und die Klausel k_3 wenden wir die Resolutions-Regel an. Zunächst haben wir

$$mgu(rot(s(d)), \neg rot(x)) = [x \mapsto s(d)]$$

Also liefert die Anwendung der Resolutions-Regel:

$$\{rot(s(d))\}, \{\neg rot(x), fliegt(x)\} \vdash \{fliegt(s(d))\}$$

5. Um die so erhaltenen Klausel $\{fliegt(s(d))\}$ mit der Klausel k_2 resolvieren zu können, berechnen wir

$$mgu(fliegt(s(d)), fliegt(s(x))) = [x \mapsto d]$$

Dann liefert die Resolutions-Regel

$$\{fliegt(s(d))\}, \{\neg fliegt(s(x)), glücklich(x)\} \vdash \{glücklich(d)\}.$$

6. Auf das Ergebnis $\{glücklich(d)\}$ und die Klausel k_6 können wir nun die Resolutions-Regel anwenden:

$$\{glücklich(d)\}, \{\neg glücklich(d)\} \vdash \{\}.$$

Da wir im letzten Schritt die leere Klausel erhalten haben, ist insgesamt $M \vdash \perp$ nachgewiesen worden und damit haben wir gezeigt, dass alle kommunistischen Drachen glücklich sind.

Aufgabe 6: Die von Bertrant Russell definierte *Russell-Menge* R ist definiert als die Menge aller der Mengen, die sich nicht selbst enthalten. Damit gilt also

$$\forall x : (x \in R \leftrightarrow \neg x \in x).$$

Zeigen Sie mit Hilfe des in diesem Abschnitt definierten Kalküls, dass diese Formel widersprüchlich ist.

Aufgabe 7: Gegeben seien folgende Axiome:

1. Jeder Barbier rasiert alle Personen, die sich nicht selbst rasieren.
2. Kein Barbier rasiert jemanden, der sich selbst rasiert.

Zeigen Sie, dass aus diesen Axiomen logisch die folgende Aussage folgt:

Alle Barbieri sind blond.

6.6 *Prover9* und *Mace4*

Der im letzten Abschnitt beschriebene Kalkül läßt sich automatisieren und bildet die Grundlage moderner automatischer Beweiser. Gleichzeitig läßt sich auch die Suche nach Gegenbeispielen automatisieren. Wir stellen in diesem Abschnitt zwei Systeme vor, die diesen Zwecken dienen.

1. *Prover9* dient dazu, automatisch prädikatenlogische Formeln zu beweisen.
2. *Mace4* untersucht, ob eine gegebene Menge prädikatenlogischer Formeln in einer endlichen Struktur erfüllbar ist. Gegebenenfalls wird diese Struktur berechnet.

Die beiden Programme *Prover9* und *Mace4* wurden von William McCune [McC10] entwickelt, stehen unter der **GPL** (*Gnu General Public Licence*) und können unter der Adresse

<http://www.cs.unm.edu/~mccune/prover9/download/>

im Quelltext heruntergeladen werden. Wir diskutieren zunächst *Prover9* und schauen uns anschließend *Mace4* an.

6.6.1 Der automatische Beweiser *Prover9*

Prover9 ist ein Programm, das als Eingabe zwei Mengen von Formeln bekommt. Die erste Menge von Formeln wird als Menge von *Axiomen* interpretiert, die zweite Menge von Formeln sind die zu beweisenden *Theoreme*, die aus den Axiomen gefolgert werden sollen. Wollen wir beispielsweise zeigen, dass in der Gruppen-Theorie aus der Existenz eines links-inversen Elements auch die Existenz eines rechts-inversen Elements folgt und dass außerdem das links-neutrale Element auch rechts-neutral ist, so können wir zunächst die Gruppen-Theorie wie folgt axiomatisieren:

1. $\forall x : e \cdot x = x,$
2. $\forall x : \exists y : x \cdot y = e,$
3. $\forall x : \forall y : \forall z : (x \cdot y) \cdot z = x \cdot (y \cdot z).$

Wir müssen nun zeigen, dass aus diesen Axiomen die beiden Formeln

$$\forall x : x \cdot e = x \quad \text{und} \quad \forall x : \exists y : y \cdot x = e$$

logisch folgen. Wir können diese Formeln wie in Abbildung 6.6 auf Seite 145 gezeigt für *Prover9* darstellen. Der Anfang der Axiome wird in dieser Datei durch “`formulas(sos)`” eingeleitet und durch das Schlüsselwort “`end_of_list`” beendet. Zu beachten ist, dass sowohl die Schlüsselwörter als auch die einzelnen Formel jeweils durch einen Punkt “.” beendet werden. Die Axiome in den Zeilen 2, 3, und 4 drücken aus, dass

1. *e* ein links-neutrales Element ist,
2. zu jedem Element *x* ein links-inverses Element *y* existiert und
3. das Assoziativ-Gesetz gilt.

Aus diesen Axiomen folgt, dass das *e* auch ein rechts-neutrales Element ist und dass außerdem zu jedem Element *x* ein rechts-neutrales Element *y* existiert. Diese beiden Formeln sind die zu beweisenden *Ziele* und werden in der Datei durch “`formulas(goal)`” markiert. Trägt die in Abbildung 6.6 gezeigte Datei den Namen “`group2.in`”, so können wir das Programm *Prover9* mit dem Befehl

```
prover9 -f group2.in
```

starten und erhalten als Ergebnis die Information, dass die beiden in Zeile 8 und 9 gezeigten Formeln tatsächlich aus den vorher angegebenen Axiomen folgen. Hätte *Prover9* keinen Beweis gefunden, so wäre das Programm solange weitergelaufen, bis kein freier Speicher mehr zur Verfügung gestanden hätte und wäre dann mit einer Fehlermeldung abgebrochen.

```

1  formulas(sos).
2  all x (e * x = x).                % left neutral
3  all x exists y (y * x = e).       % left inverse
4  all x all y all z ((x * y) * z = x * (y * z)). % associativity
5  end_of_list.
6
7  formulas(goals).
8  all x (x * e = x).                % right neutral
9  all x exists y (x * y = e).       % right inverse
10 end_of_list.

```

Abbildung 6.6: Textuelle Darstellung der Axiome der Gruppentheorie.

Prover9 versucht, einen indirekten Beweis zu führen. Zunächst werden die Axiome in prädikatenlogische Klauseln überführt. Dann wird jedes zu beweisenden Theorem negiert und die negierte Formel wird ebenfalls in Klauseln überführt. Anschließend versucht *Prover9* aus der Menge aller Axiome zusammen mit den Klauseln, die sich aus der Negation eines der zu beweisenden Theoreme ergeben, die leere Klausel herzuleiten. Gelingt dies, so ist bewiesen, dass das jeweilige Theorem tatsächlich aus den Axiomen folgt. Abbildung 6.7 zeigt eine Eingabedatei für *Prover9*, bei der versucht wird, das Kommutativ-Gesetz aus den Axiomen der Gruppentheorie zu folgern. Der Beweis-Versuch mit *Prover9* schlägt allerdings fehl. In diesem Fall wird die Beweissuche nicht endlos fortgesetzt. Dies liegt daran, dass es *Prover9* gelingt, in endlicher Zeit alle aus den gegebenen Voraussetzungen folgenden Formeln abzuleiten. Ein solcher Fall ist allerdings eher die Ausnahme als die Regel.

6.6.2 Mace4

Dauert ein Beweisversuch mit *Prover9* endlos, so ist zunächst nicht klar, ob das zu beweisende Theorem gilt. Um sicher zu sein, dass eine Formel nicht aus einer gegebenen Menge von Axiomen folgt, reicht es aus, eine Struktur zu konstruieren, in der alle Axiome erfüllt sind, in der das zu beweisende Theorem aber falsch ist. Das

```

1  formulas(sos).
2  all x (e * x = x).                % left neutral
3  all x exists y (y * x = e).        % left inverse
4  all x all y all z ((x * y) * z = x * (y * z)). % associativity
5  end_of_list.
6
7  formulas(goals).
8  all x all y (x * y = y * x).        % * is commutative
9  end_of_list.

```

Abbildung 6.7: Gilt das Kommutativ-Gesetz in allen Gruppen?

Programm *Mace4* dient genau dazu, solche Strukturen zu finden. Das funktioniert natürlich nur, solange die Strukturen endlich sind. Abbildung 6.8 zeigt eine Eingabe-Datei, mit deren Hilfe wir die Frage, ob es endliche nicht-kommutative Gruppen gibt, unter Verwendung von *Mace4* beantworten können. In den Zeilen 2, 3 und 4 stehen die Axiome der Gruppen-Theorie. Die Formel in Zeile 5 postuliert, dass für die beiden Elemente a und b das Kommutativ-Gesetz nicht gilt, dass also $a \cdot b \neq b \cdot a$ ist. Ist der in Abbildung 6.8 gezeigte Text in einer Datei mit dem Namen “group.in” gespeichert, so können wir *Mace4* durch das Kommando

mace4 -f group.in

starten. *Mace4* sucht für alle positiven natürlichen Zahlen $n = 1, 2, 3, \dots$, ob es eine Struktur $\mathcal{S} = \langle \mathcal{U}, \mathcal{J} \rangle$ mit $\text{card}(U) = n$ gibt, in der die angegebenen Formeln gelten. Bei $n = 6$ wird *Mace4* fündig und berechnet tatsächlich eine Gruppe mit 6 Elementen, in der das Kommutativ-Gesetz verletzt ist.

```

1  formulas(theory).
2  all x (e * x = x).                % left neutral
3  all x exists y (y * x = e).        % left inverse
4  all x all y all z ((x * y) * z = x * (y * z)). % associativity
5  a * b != b * a.                  % a and b do not commute
6  end_of_list.

```

Abbildung 6.8: Gibt es eine Gruppe, in der das Kommutativ-Gesetz nicht gilt?

Abbildung 6.9 zeigt einen Teil der von *Mace4* produzierten Ausgabe. Die Elemente der Gruppe sind die Zahlen $0, \dots, 5$, die Konstante a ist das Element 0, b ist das Element 1, e ist das Element 2. Weiter sehen wir, dass das Inverse von 0 wieder 0 ist, das Inverse von 1 ist 1 das Inverse von 2 ist 2, das Inverse von 3 ist 4, das Inverse von 4 ist 3 und das Inverse von 5 ist 5. Die Multiplikation wird durch die folgende Gruppen-Tafel realisiert:

\circ	0	1	2	3	4	5
0	2	3	0	1	5	4
1	4	2	1	5	0	3
2	0	1	2	3	4	5
3	5	0	3	4	2	1
4	1	5	4	2	3	0
5	3	4	5	0	1	2

Diese Gruppen-Tafel zeigt, dass

$$a \circ b = 0 \circ 1 = 3, \quad \text{aber} \quad b \circ a = 1 \circ 0 = 4$$

gilt, mithin ist das Kommutativ-Gesetz tatsächlich verletzt.

```

1  ===== DOMAIN SIZE 6 =====
2
3  === Mace4 starting on domain size 6. ===
4
5  ===== MODEL =====
6
7  interpretation( 6, [number=1, seconds=0], [
8
9      function(a, [ 0 ]),
10
11     function(b, [ 1 ]),
12
13     function(e, [ 2 ]),
14
15     function(f1(_), [ 0, 1, 2, 4, 3, 5 ]),
16
17     function(*(_,_), [
18         2, 3, 0, 1, 5, 4,
19         4, 2, 1, 5, 0, 3,
20         0, 1, 2, 3, 4, 5,
21         5, 0, 3, 4, 2, 1,
22         1, 5, 4, 2, 3, 0,
23         3, 4, 5, 0, 1, 2 ])
24 ]).
25
26  ===== end of model =====

```

Abbildung 6.9: Ausgabe von *Mace4*.

Bemerkung: Der Theorem-Beweiser *Prover9* ist ein Nachfolger des Theorem-Beweisers *Otter*. Mit Hilfe von *Otter* ist es McCune 1996 gelungen, die Robbin'sche Vermutung zu beweisen [McC97]. Dieser Beweis war damals sogar der *New York Times* eine Schlagzeile wert, nachzulesen unter

<http://www.nytimes.com/library/cyber/week/1210math.html>.

Das Ergebnis zeigt, dass automatische Theorem-Beweiser durchaus nützliche Werkzeuge sein können. Nichtdestoweniger ist die Prädikatenlogik unentscheidbar und bisher sind nur wenige offene mathematische Probleme mit Hilfe von automatischen Beweisern gelöst worden. Das wird sich vermutlich auch in der näheren Zukunft nicht ändern.

Kapitel 7

Prolog

Im diesem Kapitel wollen wir uns mit dem logischen Programmieren und der Sprache *Prolog* beschäftigen. Der Name *Prolog* steht für “*programming in logic*”. Die Grundidee des logischen Programmieren kann wie folgt dargestellt werden:

1. Der Software-Entwickler erstellt eine *Wissensbasis*. Diese enthält Informationen in Form von *Fakten* und *Regeln*.
2. Ein automatischer Beweiser (eine sogenannte *Inferenz-Maschine*) erschließt aus diesen Fakten und Regeln Informationen und beantwortet so Anfragen.

Das Besondere an dieser Art der Problemlösung besteht darin, dass es nicht mehr notwendig ist, einen Algorithmus zu entwickeln, der ein bestimmtes Problem löst. Statt dessen wird das Problem durch logische Formeln beschrieben. Zur Lösung des Problems wird dann ein automatischen Beweiser eingesetzt, der die gesuchte Lösung berechnet. Diese Vorgehensweise folgt dem Paradigma des *deklarativen Programmierens*. In der Praxis funktioniert der deklarative Ansatz nur bei einfachen Beispielen. Um mit Hilfe von *Prolog* auch komplexere Aufgaben lösen zu können, ist es unumgänglich, die Funktionsweise des eingesetzten automatischen Beweisers zu verstehen.

Wir geben ein einfaches Beispiel, an dem wir das Grundprinzip des deklarativen Programmierens mit *Prolog* erläutern können. Abbildung 7.1 auf Seite 149 zeigt ein *Prolog*-Programm, das aus einer Ansammlung von *Fakten* und *Regeln* besteht:

1. Ein *Fakt* ist eine atomare Formel. Die Syntax ist

$$p(t_1, \dots, t_n).$$

Dabei ist p ein Prädikats-Zeichen und t_1, \dots, t_n sind Terme. Ist die Menge der Variablen, die in den Termen t_1, \dots, t_n vorkommen, durch $\{x_1, \dots, x_m\}$ gegeben, so wird der obige Fakt als die logische Formel

$$\forall x_1, \dots, x_m: p(t_1, \dots, t_n)$$

interpretiert. Das Programm in Abbildung 7.1 enthält in den Zeilen 1 bis 5 Fakten. Umgangssprachlich können wir diese wie folgt lesen:

- (a) Asterix ist ein Gallier.
- (b) Obelix ist ein Gallier.
- (c) Cäsar ist ein Kaiser.
- (d) Cäsar ist ein Römer.

2. Eine *Regel* ist eine bedingte Aussage. Die Syntax ist

$$A :- B_1, \dots, B_n.$$

Dabei sind A und B_1, \dots, B_n atomare Formeln, haben also die Gestalt

$$q(s_1, \dots, s_k),$$

wobei q ein Prädikats-Zeichen ist und s_1, \dots, s_k Terme sind. Ist die Menge der Variablen, die in den atomaren Formeln A, B_1, \dots, B_n auftreten, durch $\{x_1, \dots, x_m\}$ gegeben, so wird die obige Regel als die logische Formel

$$\forall x_1, \dots, x_m : (B_1 \wedge \dots \wedge B_m \rightarrow A)$$

interpretiert. Das Programm aus Abbildung 7.1 enthält in den Zeilen 7 bis 12 Regeln. Schreiben wir diese Regeln als prädikatenlogische Formeln, so erhalten wir:

- (a) $\forall x : (\text{gallier}(x) \rightarrow \text{stark}(x))$
Alle Gallier sind stark.
- (b) $\forall x : (\text{stark}(x) \rightarrow \text{maechtig}(x))$
Wer stark ist, ist mächtig.
- (c) $\forall x : (\text{kaiser}(x) \wedge \text{roemer}(x) \rightarrow \text{maechtig}(x))$
Wer Kaiser und Römer ist, der ist mächtig.
- (d) $\forall x : (\text{roemer}(x) \rightarrow \text{spinnt}(x))$
Wer Römer ist, spinnt.

```

1  gallier(asterix).
2  gallier(obelix).
3
4  kaiser(caesar).
5  roemer(caesar).
6
7  stark(X) :- gallier(X).
8
9  maechtig(X) :- stark(X).
10 maechtig(X) :- kaiser(X), roemer(X).
11
12 spinnt(X) :- roemer(X).
```

Abbildung 7.1: Ein einfaches Prolog-Programm.

Wir haben in dem Programm in Abbildung 7.1 die Zeile

`stark(X) :- gallier(X).`

als die Formel

$$\forall x : (\text{gallier}(x) \rightarrow \text{stark}(x))$$

interpretiert. Damit eine solche Interpretation möglich ist, muss klar sein, dass in der obigen Regel der String “X” eine Variable bezeichnet, während die Strings “stark” und “gallier” Prädikats-Zeichen sind. Prolog hat sehr einfache syntaktische Regeln um Variablen von Prädikats- und Funktions-Zeichen unterscheiden zu können:

1. Wenn ein String mit einem großen Buchstaben oder aber mit dem Unterstrich “_” beginnt und nur aus Buchstaben, Ziffern und dem Unterstrich “_” besteht, dann bezeichnet dieser String eine Variable. Die folgenden Strings sind daher Variablen:

X, ABC_32, _U, Hugo, _1, _.

2. Strings, die mit einem kleinen Buchstaben beginnen und nur aus Buchstaben, Ziffern und dem Unterstrich “_” bestehen, bezeichnen Prädikats- und Funktions-Zeichen. Die folgenden Strings können also als Funktions- oder Prädikats-Zeichen verwendet werden:

`asterix, a1, i_love_prolog, x.`

3. Die Strings

`“+”, “-”, “*”, “/”, “.”`

bezeichnen Funktions-Zeichen. Bis auf das Funktions-Zeichen “.” können diese Funktions-Zeichen auch, wie in der Mathematik üblich, als Infix-Operatoren geschrieben werden. Das Funktions-Zeichen “.” bezeichnen wir als *Dot-Operator*. Dieser Operator wird zur Konstruktion von Listen benutzt. Die Details werden wir später diskutieren.

4. Die Strings

`“<”, “>”, “=”, “=<”, “>=”, “\=”, “==”, “\==”.`

bezeichnen Prädikats-Zeichen. Auch für diese Prädikats-Zeichen ist eine Infix-Schreibweise zulässig.

Gegenüber den Sprachen C und SETLX gibt es hier die folgenden Unterschiede, die besonders Anfängern oft Probleme bereiten.

- Bei dem Operator “=<” treten die Zeichen “=” und “<” nicht in der selben Reihenfolge auf, wie das in den Sprachen C oder SETLX der Fall ist, denn dort wird dieser Operator als “<=” geschrieben.
- Der Operator “==” testet, ob die beiden Argumente gleich sind, während der Operator “\==” testet, ob die beiden Werte ungleich sind. In C und SETLX hat der entsprechende Operator die Form “!=”.
- Der Operator “=” ist der Unifikations-Operator. Bei einem Aufruf der Form

$$s = t$$

versucht das *Prolog*-System die syntaktische Gleichung $s \doteq t$ zu lösen.¹ Falls dies erfolgreich ist, werden die Variablen, die in den Termen s und t vorkommen, *instantiiert*. Was dabei im Detail passiert, werden wir später sehen.

Der Operator “\=” bezeichnet die Negation des Unifikations-Operators.

5. Zahlen bezeichnen 0-stellige Funktions-Zeichen. In *Prolog* können Sie sowohl ganze Zahlen als auch Fließkomma-Zahlen benutzen. Die Syntax für Zahlen ist ähnlich wie in C, die folgenden Strings stellen also Zahlen dar:

`12, -3, 2.5, 2.3e-5.`

Nachdem wir jetzt Syntax und Semantik des Programms in Abbildung 7.1 erläutert haben, zeigen wir nun, wie *Prolog* sogenannte *Anfragen* beantwortet. Wir nehmen an, dass das Programm in einer Datei mit dem Namen “`gallier.pl`” abgespeichert ist. Wir wollen herausfinden, ob es jemanden gibt, der einerseits mächtig ist und der andererseits spinnt. Logisch wird dies durch die folgende Formel ausgedrückt:

$$\exists x: (\text{maechtig}(x) \wedge \text{spinnt}(x)).$$

Als *Prolog*-Anfrage können wir den Sachverhalt wie folgt formulieren:

`maechtig(X), spinnt(X).`

Um diese Anfrage auswerten zu können, starten wir das *SWI-Prolog*-System² in einer Shell mit dem Kommando:

`swipl`

Wenn das *Prolog*-System installiert ist, begrüßt uns das System wie folgt:

¹ Leider wird dabei im Normalfall der *Occur-Check* nicht durchgeführt. Durch Eingabe des Befehls

`set_prolog_flag(occurs_check, true).`

können Sie aber erzwingen, dass bei jeder Unifikation ein Occur-Check durchgeführt wird.

² Sie finden dieses Prolog-System im Netz unter www.swi-prolog.org.

```
Welcome to SWI-Prolog (Multi-threaded, 64 bits, Version 6.6.1-DIRTY)
Copyright (c) 1990-2013 University of Amsterdam, VU Amsterdam
SWI-Prolog comes with ABSOLUTELY NO WARRANTY. This is free software,
and you are welcome to redistribute it under certain conditions.
Please visit http://www.swi-prolog.org for details.
```

```
For help, use ?- help(Topic). or ?- apropos(Word).
```

```
?-
```

Die Zeichenfolge “?-” ist der *Prolog*-Prompt. Hier geben wir

```
consult(gallier).
```

ein (und zwar mit dem Punkt) und drücken *Return*. Damit fordern wir das System auf, die Fakten und Regeln aus der Datei “gallier.pl” zu laden. Als Ergebnis erhalten wir die Meldung

```
?- consult(gallier).
% gallier compiled 0.00 sec, 1,676 bytes
```

```
Yes
```

```
?-
```

Das Programm wurde erfolgreich geladen und übersetzt. Wir geben nun unsere Anfrage ein

```
?- maechtig(X), spinnt(X).
```

und erhalten als Antwort:

```
X = caesar.
```

Da diese Antwort mit einem Punkt “.” beendet wird, können wir folgern, dass es keine weiteren möglichen Antworten gibt, die logisch aus den Fakten und Regeln folgen. Würde es noch weitere Antworten geben, so hätte die Antwort nur die Form

```
X = caesar
```

gehabt. In diesem Fall hätten wir zwei Möglichkeiten gehabt:

1. Wenn wir mit dieser Antwort zufrieden sind, drücken wir *Return* und erhalten einen neuen Prompt.
2. Wenn wir statt dessen nach weiteren Personen suchen wollen, die einerseits mächtig sind und andererseits spinnen, dann geben wir das Zeichen “;” ein. In diesem Fall hätte das System versucht, weitere mögliche Antworten zu finden.

Wenn wir das *Prolog*-System wieder verlassen wollen, dann geben wir den Befehl “halt.” ein.

7.1 Wie arbeitet *Prolog*?

Das Konzept des logischen Programmierens sieht vor, dass der Benutzer eine Datenbank mit Fakten und Regeln erstellt, die das Problem vollständig beschreiben. Anschließend beantwortet dann das *Prolog*-System mögliche Anfragen mit Hilfe einer Inferenz-Maschine. Um nicht-triviale *Prolog*-Programme erstellen zu können, ist es notwendig zu verstehen, wie das *Prolog*-System Anfragen beantwortet. Um diesen Algorithmus leichter darstellen zu können, vereinbaren wir folgendes: Ist ein Fakt der Form

```
A.
```

gegeben, so formen wir dies zu der Regel

```
A :- true.
```


um. Außerdem bezeichnen wir bei einer Klausel

$$A :- B_1, \dots, B_n.$$

die atomare Formel A als den *Kopf* und die Konjunktion B_1, \dots, B_n als den *Rumpf* der Klausel.

Wir beschreiben nun den Algorithmus, mit dem das *Prolog*-System Anfragen beantwortet.

1. Gegeben

(a) Anfrage: $G = Q_1, \dots, Q_n$

Hier sind Q_1, \dots, Q_n atomare Formeln.

(b) *Prolog*-Programm: P

Da die Reihenfolge der Klauseln für das Folgende relevant ist, fassen wir das *Prolog*-Programm P als Liste von Regeln auf.

2. **Gesucht:** Eine Substitution σ , so dass die Instanz $G\sigma$ aus den Regeln des Programms P folgt:

$$\models \forall(P) \rightarrow \forall(G\sigma).$$

Hier bezeichnet $\forall(P)$ den Allabschluß der Konjunktion aller Klauseln aus P und $\forall(G\sigma)$ bezeichnet entsprechend den Allabschluß von $G\sigma$.

Der Algorithmus selbst arbeitet wie folgt:

1. Suche (der Reihe nach) in dem Programm P alle Regeln

$$A :- B_1, \dots, B_m.$$

für die der Unifikator $\mu = \text{mgu}(Q_1, A)$ existiert.

2. Gibt es mehrere solche Regeln, so

(a) wählen wir die erste Regel aus, wobei wir uns an der Reihenfolge orientieren, in der die Regeln in dem Programm P auftreten.

(b) Außerdem setzen wir an dieser Stelle einen Auswahl-Punkt (*Choice-Point*), um später hier eine andere Regel wählen zu können, falls dies notwendig werden sollte.

3. Wir setzen $G := G\mu$, wobei μ der oben berechnete Unifikator ist.

4. Nun bilden wir die Anfrage

$$B_1\mu, \dots, B_m\mu, Q_2\mu, \dots, Q_n\mu.$$

Jetzt können zwei Fälle auftreten:

(a) $m + n = 0$: Dann ist die Beantwortung der Anfrage erfolgreich und wir geben als Antwort $G\mu$ zurück.

(b) Sonst beantworten wir rekursiv die Anfrage $B_1\mu, \dots, B_m\mu, Q_2\mu, \dots, Q_n\mu$.

Falls die rekursive Beantwortung unter Punkt 4 erfolglos war, gehen wir zum letzten Auswahl-Punkt zurück. Gleichzeitig werden alle Zuweisungen $G := G\mu$, die wir seit diesem Auswahl-Punkt durchgeführt haben, wieder rückgängig gemacht. Anschließend versuchen wir, mit der nächsten möglichen Regel die Anfrage zu beantworten.

Um den Algorithmus besser zu verstehen, beobachten wir die Abarbeitung der Anfrage

$$\text{maechtig}(X), \text{spinnt}(X).$$

im Debugger des *Prolog*-Systems. Wir geben dazu nach dem Laden des Programms "`gallier.pl`" das Kommando `guitracer` ein.

```
?- guitracer.
```

Wir erhalten die Antwort:

```
% The graphical front-end will be used for subsequent tracing
```


```
Yes
```

```
?-
```


Jetzt starten wir die ursprüngliche Anfrage noch einmal, setzen aber das Kommando `trace` vor unsere Anfrage:

```
?- trace, maechtig(X), spinnt(X).
```


Als Ergebnis wird ein Fenster geöffnet, das Sie in Abbildung 7.2 auf Seite 154 sehen. Unter dem Menü sehen Sie hier eine Werkzeugleiste. Die einzelnen Symbole haben dabei die folgende Bedeutung:

1. Die Schaltfläche  dient dazu, einzelne Unifikationen anzuzeigen. Mit dieser Schaltfläche können wir die meisten Details der Abarbeitung beobachten.

Alternativ hat die Taste “i” die selbe Funktion.

2. Die Schaltfläche  dient dazu, einen einzelnen Schritt bei der Abarbeitung einer Anfrage durchzuführen.


Alternativ hat die Leertaste die selbe Funktion.

3. Die Schaltfläche  dient dazu, die nächste atomare Anfrage in einem Schritt durchzuführen. Drücken wir diese Schaltfläche unmittelbar, nachdem wir das Ziel

```
maechtig(X), roemer(X).
```

einggegeben haben, so würde der Debugger die Anfrage `maechtig(X)` in einem Schritt beantworten. Dabei würde dann die Variable `X` an die Konstante `asterix` gebunden.


Alternativ hat die Taste “s” (*skip*) die selbe Funktion.

4. Die Schaltfläche  dient dazu, die Prozedur, in deren Abarbeitung wir uns befinden, ohne Unterbrechung zu Ende abarbeiten zu lassen. Versuchen wir beispielsweise die atomare Anfrage “`maechtig(X)`” mit der Regel

```
maechtig(X) :- kaiser(X), roemer(X).
```

abzuarbeiten und müssen nun die Anfrage “`kaiser(X), roemer(X).`” beantworten, so würden wir durch Betätigung dieser Schaltfläche sofort die Antwort “`X = caesar`” für diese Anfrage erhalten.

Alternative hat die Taste “f” (*finish*) die selbe Funktion.



5. Die Schaltfläche  dient dazu, eine bereits beantwortete Anfrage noch einmal zu beantworten. Dies kann sinnvoll sein, wenn man beim Tracen einer Anfrage nicht aufgepaßt hat und noch einmal sehen möchte, wie das *Prolog*-System zu seiner Antwort gekommen ist.

Alternative hat die Taste “r” (*retry*) die selbe Funktion.

6. Die Schaltfläche  beantwortet die gestellte Anfrage ohne weitere Unterbrechung.

Alternative hat die Taste “n” (*nodebug*) die selbe Funktion.

Von den weiteren Schaltflächen sind fürs erste nur noch zwei interessant.

7. Die Schaltfläche  läßt das Programm bis zum nächsten Haltepunkt laufen.
Alternative hat die Taste “1” (*continue*) die selbe Funktion.
8. Die Schaltfläche  setzt einen Haltepunkt. Dazu muss vorher der Cursor an die Stelle gebracht werden, an der ein Haltepunkt gesetzt werden soll.
Alternative hat die Taste “!” die selbe Funktion.

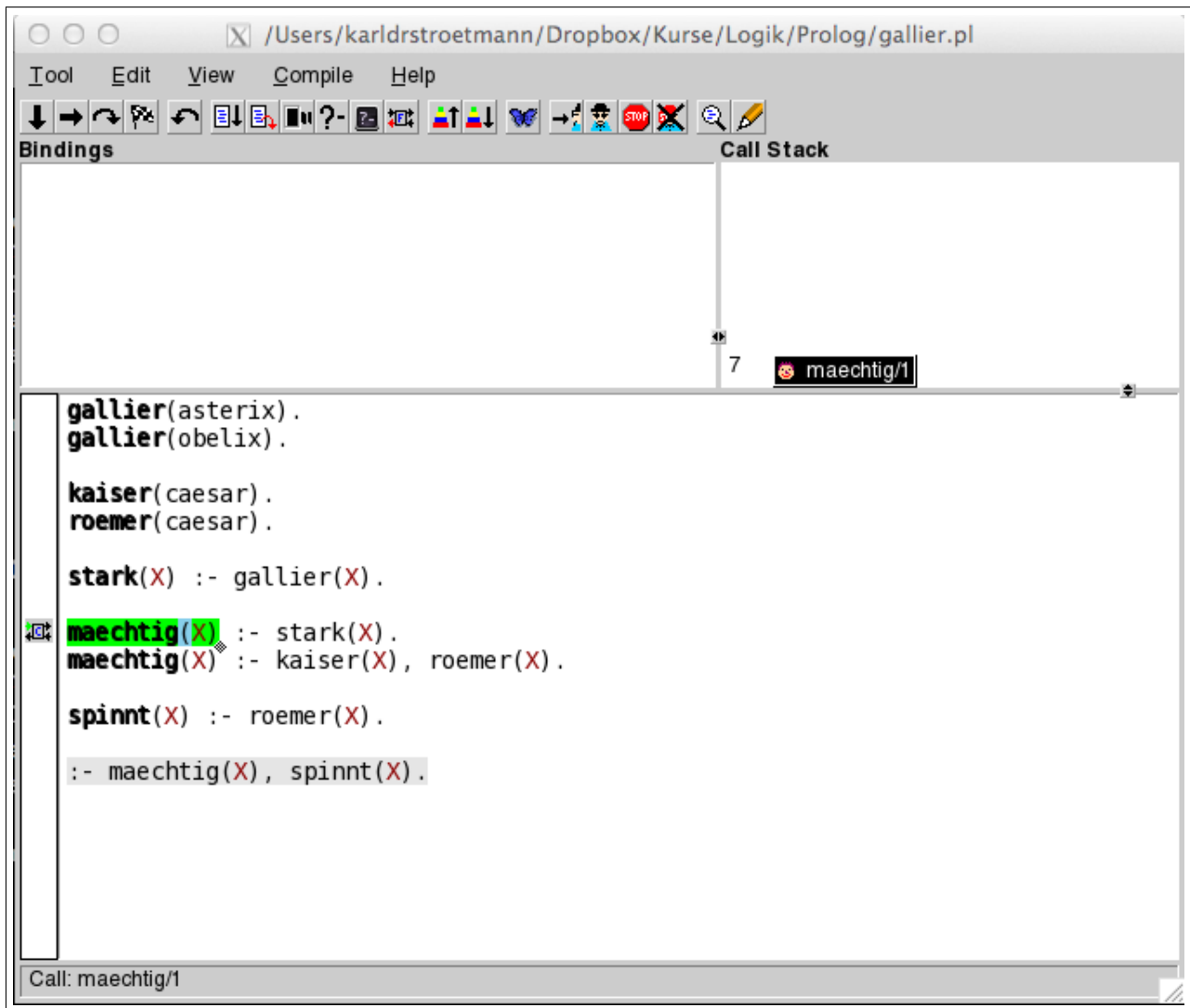


Abbildung 7.2: Der Debugger des *SWI-Prolog*-Systems.

Wir zeigen, wie das Ziel `maechtig(X)`, `spinnt(X)` vom *Prolog*-System beantwortet wird.

1. Zunächst wird versucht, die Anfrage “`maechtig(X)`” zu lösen. Die erste Regel, die das Prädikat `maechtig/1` definiert, ist

`maechtig(X) :- stark(X).`

Daher wird die Anfrage “`maechtig(X)`” reduziert zu der Anfrage “`stark(X)`”. Die aktuelle vollständige Anfrage lautet nun

`stark(X), spinnt(X).`

Da es noch eine zweite Regel gibt, die das Prädikat `maechtig/1` definiert, setzen wir an dieser Stelle einen Auswahl-Punkt (*Choice-Point*). Falls also die Beantwortung der Anfrage “`stark(X), spinnt(X)`” später scheitert, können wir es mit der zweiten Regel noch einmal versuchen.

2. Jetzt wird versucht, die Anfrage “`stark(X)`” zu lösen. Die erste und einzige Regel, die das Prädikat `stark/1` definiert, ist

`stark(X) :- gallier(X).`

Nach der Unifikation des Kopfes dieser Regel mit der Anfrage “`stark(X)`” lautet die aktuelle Anfrage

`gallier(X), spinnt(X).`

3. Die erste Regel, die das Prädikat `gallier/1` definiert und deren Kopf mit der Anfrage “`gallier(X)`” unifiziert werden kann, ist der Fakt

`gallier(asterix).`

Bei der Unifikation mit diesem Fakt wird die Variable `X` an die Konstante `asterix` gebunden. Damit lautet jetzt die aktuelle Anfrage

`spinnt(asterix).`

Da es noch eine zweite Regel gibt, die das Prädikat `gallier/1` definiert, setzen wir an dieser Stelle einen Auswahl-Punkt.

4. Die erste und einzige Regel, die das Prädikat `spinnt/1` definiert, lautet

`spinnt(X) :- roemer(X).`

Also wird nun die Variable `X` in dieser Regel mit `asterix` unifiziert und wir erhalten die Anfrage

`roemer(asterix).`

5. Die einzige Regel, die das Prädikat `roemer/1` definiert, ist

`roemer(caesar).`

Diese Regel läßt sich nicht mit der Anfrage “`roemer(asterix)`” unifizieren. Also scheitert diese Anfrage.

6. Wir schauen nun, wann wir das letzte mal einen Auswahl-Punkt gesetzt haben. Wir stellen fest, dass wir unter Punkt 3 bei der Beantwortung der Anfrage `gallier(X)` das letzte Mal einen Auswahl-Punkt gesetzt haben. Also gehen wir nun zu Punkt 3 zurück und versuchen wieder, die Anfrage

`gallier(X), spinnt(X)`

zu lösen. Diesmal wählen wir jedoch den Fakt

`gallier(obelix).`

Wir erhalten dann die neue Anfrage

`spinnt(obelix).`

7. Die erste und einzige Regel, die das Prädikat `spinnt/1` definiert, lautet

`spinnt(X) :- roemer(X).`

Also wird die Variable `X` in dieser Regel mit `obelix` unifiziert und wir erhalten die Anfrage

`roemer(obelix).`

8. Die einzige Regel, die das Prädikat `roemer/1` definiert, ist

`roemer(caesar).`

Diese Regel läßt sich nicht mit der Anfrage “`roemer(asterix)`” unifizieren. Also scheitert diese Anfrage.

9. Wir schauen wieder, wann das letzte Mal ein Auswahl-Punkt gesetzt wurde. Der unter Punkt 3 gesetzte Auswahl-Punkt wurde vollständig abgearbeitet, dieser Auswahl-Punkt kann uns also nicht mehr helfen. Aber unter Punkt 1 wurde ebenfalls ein Auswahl-Punkt gesetzt, denn für das Prädikat `maechtig/1` gibt es die weitere Regel

```
maechtig(X) :- kaiser(X), roemer(X).
```

Wenden wir diese Regel an, so erhalten wir die Anfrage

```
kaiser(X), roemer(X), spinnt(X).
```

10. Für das Prädikat `kaiser/1` enthält unsere Datenbank genau einen Fakt:

```
kaiser(caesar).
```

Benutzen wir diesen Fakt zur Reduktion unserer Anfrage, so lautet die neue Anfrage

```
roemer(caesar), spinnt(caesar).
```

11. Für das Prädikat `roemer/1` enthält unsere Datenbank genau einen Fakt:

```
roemer(caesar).
```

Benutzen wir diesen Fakt zur Reduktion unserer Anfrage, so lautet die neue Anfrage

```
spinnt(caesar).
```

12. Die erste und einzige Regel, die das Prädikat `spinnt/1` definiert, lautet

```
spinnt(X) :- roemer(X).
```

Also wird die Variable `X` in dieser Regel mit `caesar` unifiziert und wir erhalten die Anfrage

```
roemer(caesar).
```

13. Für das Prädikat `roemer/1` enthält unsere Datenbank genau einen Fakt:

```
roemer(caesar).
```

Benutzen wir diesen Fakt zur Reduktion unserer Anfrage, so ist die verbleibende Anfrage leer. Damit ist die ursprüngliche Anfrage gelöst. Die dabei berechnete Antwort erhalten wir, wenn wir untersuchen, wie die Variable `X` unifiziert worden ist. Die Variable `X` war unter Punkt 10 mit der Konstanten `caesar` unifiziert worden. Also ist

```
X = caesar
```

die Antwort, die von dem System berechnet wird.

Bei der Beantwortung der Anfrage "`maechtig(X), spinnt(X)`" sind wir einige Male in Sackgassen hineingelaufen und mussten Instantiierungen der Variable `X` wieder zurück nehmen. Dieser Vorgang wird in der Literatur als *backtracking* bezeichnet. Er kann mit Hilfe des Debuggers am Bildschirm verfolgt werden.

7.1.1 Die Tiefensuche

Der von Prolog verwendete Such-Algorithmus wird auch als *Tiefensuche* (angelsächsisch: *depth first search*) bezeichnet. Um diesen Ausdruck erläutern zu können, definieren wir zunächst den Begriff des *Suchbaums*. Die Knoten eines Suchbaums sind mit Anfrage beschriftet. Ist der Knoten u des Suchbaums mit der Anfrage

$$Q_1, \dots, Q_m$$

beschriftet und gibt es für $i = 1, \dots, k$ Regeln der Form

$$A^{(i)} :- B_1^{(i)}, \dots, B_{n(i)}^{(i)}$$

die auf die Anfrage passen, für die also $\mu_i = mgu(Q_1, A^{(i)})$ existiert, so hat der Knoten u insgesamt k verschiedene Kinder. Dabei ist das i -te Kind mit der Anfrage

$$B_1^{(i)}\mu_i, \dots, B_{n(i)}^{(i)}\mu_i, Q_2\mu_i, \dots, Q_m\mu_i$$

beschriftet. Als einfaches Beispiel betrachten wir das in Abbildung 7.3 gezeigte Programm. Der Suchbaum für die Anfrage

$$p(X)$$

ist in Abbildung 7.4 gezeigt. Den Knoten, der mit der ursprünglichen Anfrage beschriftet ist, bezeichnen wir als die *Wurzel* des Suchbaums. Die *Lösungen* zu der ursprünglichen Anfrage finden wir an den *Blättern* des Baumes: Wir bezeichnen hier die Knoten als Blätter, die am weitesten unten stehen. Suchbäume stehen also gewissermaßen auf dem Kopf: Die Blätter sind unten und die Wurzeln sind oben³.

```

1  p(X) :- q1(X).
2  p(X) :- q2(X).
3
4  q1(X) :- r1(X).
5  q1(X) :- r2(X).
6
7  q2(X) :- r3(X).
8  q2(X) :- r4(X).
9
10 r1(a).
11 r2(b).
12 r3(c).
13 r4(d).
```

Abbildung 7.3: Die Tiefensuche in Prolog

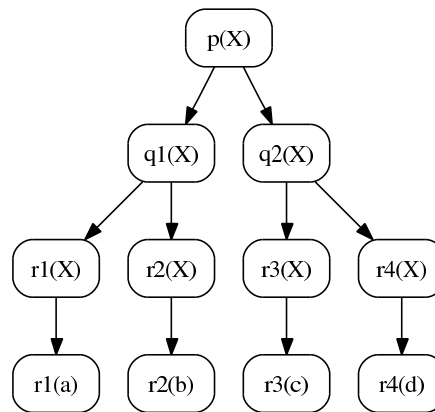


Abbildung 7.4: Der Suchbaum für das in Abbildung 7.3 gezeigte Programm.

Anhand des in Abbildung 7.4 gezeigten Suchbaums läßt sich nun die Tiefensuche erklären: Wenn der Prolog-Interpreter nach einer Lösung sucht, so wählt er immer den linken Ast und steigt dort so tief wie möglich ab. Dadurch werden die Lösungen in dem Beispiel in der Reihenfolge

$$X = a, X = b, X = c, X = d,$$

³ Daher werden diese Suchbäume auch als australische Bäume bezeichnet.

gefunden. Die Tiefensuche ist dann problematisch, wenn der linke Ast des Suchbaums unendlich tief ist. Als Beispiel betrachten wir das in Abbildung 7.5 gezeigte Prolog-Programm. Zeichnen wir hier den Suchbaum für die Anfrage

$p(X)$

so finden wir einen unendlichen Ast, in den der Prolog-Interpreter absteigt und aus dem er dann mit einem Stack-Overflow wieder zurück kommt. Vertauschen wir hingegen die Reihenfolge der Klauseln, so kann das Programm die obige Anfrage beantworten.

```

1  p(s(X)) :- p(X).
2  p(c).
```

Abbildung 7.5: Eine Endlos-Schleife in *Prolog*.

7.2 Ein komplexeres Beispiel

Das obige Beispiel war bewußt einfach gehalten um die Sprache *Prolog* einzuführen. Um die Mächtigkeit des Backtrackings zu demonstrieren, präsentieren wir jetzt ein komplexeres Beispiel. Es handelt sich um das folgende Rätsel:

1. Drei Freunde belegen den ersten, zweiten und dritten Platz bei einem Programmier-Wettbewerb.
2. Jeder der drei hat genau einen Vornamen, genau ein Auto und hat sein Programm in genau einer Programmier-Sprache geschrieben.
3. Michael programmiert in SETLX und war besser als der Audi-Fahrer.
4. Julia, die einen Ford Mustang fährt, war besser als der Java-Programmierer.
5. Das Prolog-Programm war am besten.
6. Wer fährt Toyota?
7. In welcher Sprache programmiert Thomas?

Um dieses Rätsel zu lösen, überlegen wir uns zunächst, wie wir die einzelnen Daten repräsentieren können, die in dem Rätsel eine Rolle spielen. Zunächst ist dort von Personen die Rede. Jede dieser Personen hat genau einen Vornamen, ein Auto und eine Programmier-Sprache. Wir repräsentieren Personen daher durch Terme der Form

`person(Name, Car, Language).`

Dabei bezeichnen *Name*, *Car* und *Language* Konstanten, die aus den entsprechenden Mengen gewählt werden:

$Name \in \{\text{ julia, thomas, michael} \}, \quad Car \in \{\text{ ford, toyota, audi} \},$

$Language \in \{\text{ java, prolog, setlX} \}.$

Wenn wir Personen durch ein dreistelliges Funktions-Zeichen wie oben gezeigt repräsentieren, können wir sofort Prädikate angeben, die den Vornamen, die Auto-Marke und die Programmier-Sprache aus einem solchen Term extrahieren.

1. Das Prädikat `first_name/2` extrahiert den Vornamen:

`first_name(person(Name, Car, Language), Name).`

2. Das Prädikat `car/2` extrahiert die Auto-Marke:

`car(person(Name, Car, Language), Car).`

3. Das Prädikat `language/2` extrahiert die Programmier-Sprache:

```
language(person(Name, Car, Language), Language).
```

Um zu verstehen wie diese Prädikate arbeiten, zeigen wir, wie die Anfrage

```
car( person(hans, seat, setlX), X ).
```

von dem *Prolog*-System beantwortet wird. Die einzige Regel, die zur Beantwortung dieser Anfrage herangezogen werden kann, ist die Regel

```
car(person(Name, Car, Language), Car) :- true.
```

Um diese Regel anwenden zu können, ist die syntaktische Gleichung

```
car( person(hans, seat, setlX), X )  $\doteq$  car( person(Name, Car, Language), Car )
```

zu lösen. Bei der Unifikation findet sich die Lösung

```
 $\mu = [\text{Name} \mapsto \text{hans}, \text{Car} \mapsto \text{seat}, \text{Language} \mapsto \text{setlX}, \text{X} \mapsto \text{seat}]$ .
```

Insbesondere wird also die Variable `X` bei dieser Anfrage an die Konstante `seat` gebunden.

Wie können wir nun die Reihenfolge repräsentieren, in der die drei Personen bei dem Wettbewerb abgeschnitten haben? Wir wählen ein dreistelliges Funktions-Zeichen `sequence` und repräsentieren die Reihenfolge durch den Term

```
sequence(First, Second, Third).
```

Dabei stehen *First*, *Second* und *Third* für Terme, die von dem Funktions-Zeichen `person/3` erzeugt worden sind und die Personen bezeichnen. Die Reihenfolge kann dann durch das Prädikat

```
did.better(Better, Worse, Sequence)
```

berechnet werden, dessen Implementierung in den Zeilen 38 – 40 der Abbildung 7.6 auf Seite 161 gezeigt ist. Wir können nun daran gehen, das Rätsel zu lösen. Abbildung 7.6 zeigt die Implementierung. Zeile 1 – 30 enthält die Implementierung einer Regel für das Prädikat `answer/2`, dass die Lösung des Rätsel berechnet. In dieser Regel haben wir das Rätsel als prädikatenlogische Formel codiert. Wir übersetzen diese Regel jetzt zurück in die Umgangssprache und zeigen dadurch, dass das Prädikat `answer/2` das Rätsel korrekt beschreibt. Die Numerierung in der folgenden Aufzählung stimmt jeweils mit der entsprechenden Zeilen-Nummer im Programm überein:

2. Falls `Sequence` eine Reihenfolge von drei Personen beschreibt und
4. `Michael` eine Person aus dieser Reihenfolge ist und
5. der Name der durch `Michael` bezeichneten Person den Wert `michael` hat und
6. `Michael` in `SETLX` programmiert und
8. `Audi` eine Person aus der Reihenfolge `Sequence` ist und
9. `Michael` beim Wettbewerb besser abgeschnitten hat als die durch `Audi` bezeichnete Person
10. die durch `Audi` bezeichnete Person einen Audi fährt und
- ⋮
24. `Toyota` eine Person aus der Reihenfolge `Sequence` ist und
25. die durch `Toyota` bezeichnete Person einen Toyota fährt und
26. `NameToyota` den Vornamen der durch `Toyota` bezeichneten Person angibt und
28. `Thomas` eine Person aus der Reihenfolge `Sequence` ist und
25. die durch `Thomas` bezeichnete Person den Vornamen Thomas hat und

26. `LanguageThomas` die Sprache ist, in der die durch `Thomas` bezeichnete Person programmiert, dann gilt:

1. `NameToyota` ist der Namen des Toyota-Fahrers und `LanguageThomas` ist die Sprache, in der Thomas programmiert.

Wenn wir die ursprüngliche Aufgabe mit der Implementierung in *Prolog* vergleichen, dann stellen wir fest, dass die in dem Rätsel gemachten Angaben eins-zu-eins in *Prolog* übersetzt werden konnten. Diese Übersetzung beschreibt nur das Rätsel und gibt keinen Hinweis, wie dieses Rätsel zu lösen ist. Für die Lösung ist dann die dem *Prolog*-System zu Grunde liegende *Inferenz-Maschine* zuständig.

7.3 Listen

In Prolog wird viel mit Listen gearbeitet. Listen werden in Prolog mit dem 2-stelligen Funktions-Zeichen “.” konstruiert. Ein Term der Form

`.(s,t)`

steht also für eine Liste, die als erstes Element “s” enthält. “t” bezeichnet den Rest der Liste. Das Funktions-Zeichen “[]” steht für die leere Liste. Eine Liste, die aus den drei Elementen “a”, “b” und “c” besteht, kann also wie folgt dargestellt werden:

`.(a, .(b, .(c, [])))`

Da dies relativ schwer zu lesen ist, darf diese Liste auch als

`[a,b,c]`

geschrieben werden. Zusätzlich kann der Term “.(s,t)” in der Form

`[s | t]`

geschrieben werden. Um diese Kurzschreibweise zu erläutern, geben wir ein kurzes Prolog-Programm an, das zwei Listen aneinander hängen kann. Das Programm implementiert das dreistellige Prädikat `myAppend`⁴. Die Intention ist, dass `myAppend(l1, l2, l3)` für drei Listen *l₁*, *l₂* und *l₃* genau dann wahr sein soll, wenn die Liste *l₃* dadurch entsteht, dass die Liste *l₂* hinten an die Liste *l₁* angehängt wird. Das Programm besteht aus den folgenden beiden Klauseln:

```
myAppend( [], L, L ).
myAppend( [ X | L1 ], L2, [ X | L3 ] ) :- myAppend( L1, L2, L3 ).
```

Wir können diese beiden Klauseln folgendermaßen in die Umgangssprache übersetzen:

1. Hängen wir eine Liste *L* an die leere Liste an, so ist das Ergebnis die Liste *L*.
2. Um an eine Liste `[X | L1]`, die aus dem Element *X* und dem Rest *L1* besteht, eine Liste *L2* anzuhängen, hängen wir zunächst an die Liste *L1* die Liste *L2* an und nennen das Ergebnis *L3*. Das Endergebnis erhalten wir, wenn wir vor die Liste *L3* noch das Element *X* setzen. Wir erhalten dann die Liste `[X | L3]`.

Wir testen unser Programm und nehmen dazu an, dass die beiden Programm-Klauseln in der Datei “`myAppend.pl`” abgespeichert sind und dass wir diese Datei mit dem Befehl “`consult(myAppend).`” geladen haben. Dann stellen wir die Anfrage

```
?- myAppend( [ 1, 2, 3 ], [ a, b, c ], L ).
```

Wir erhalten die Antwort:

```
L = [1, 2, 3, a, b, c]
```

⁴In dem *SWI-Prolog*-System gibt es das vordefinierte Prädikat `append/3`, das genau das selbe leistet wie unsere Implementierung von `myAppend/3`.

```

1  answer(NameToyota, LanguageThomas) :-
2      is_sequence( Sequence ),
3      % Michael programmiert in SetlX.
4      one_of_them(Michael, Sequence),
5      first_name(Michael, michael),
6      language(Michael, setlX),
7      % Michael war besser als der Audi-Fahrer
8      one_of_them(Audi, Sequence),
9      did_better(Michael, Audi, Sequence),
10     car(Audi, audi),
11     % Julia fährt einen Ford Mustang.
12     one_of_them(Julia, Sequence),
13     first_name(Julia, julia),
14     car(Julia, ford),
15     % Julia war besser als der Java-Programmierer.
16     one_of_them(JavaProgrammer, Sequence),
17     language(JavaProgrammer, java),
18     did_better(Julia, JavaProgrammer, Sequence),
19     % Das Prolog-Programm war am besten.
20     one_of_them(PrologProgrammer, Sequence),
21     first(PrologProgrammer, Sequence),
22     language(PrologProgrammer, prolog),
23     % Wer fährt Toyota?
24     one_of_them(Toyota, Sequence),
25     car(Toyota, toyota),
26     first_name(Toyota, NameToyota),
27     % In welcher Sprache programmiert Thomas?
28     one_of_them(Thomas, Sequence),
29     first_name(Thomas, thomas),
30     language(Thomas, LanguageThomas).
31
32 is_sequence( sequence(_First, _Second, _Third) ).
33
34 one_of_them(A, sequence(A, _, _)).
35 one_of_them(B, sequence(_, B, _)).
36 one_of_them(C, sequence(_, _, C)).
37
38 did_better(A, B, sequence(A, B, _)).
39 did_better(A, C, sequence(A, _, C)).
40 did_better(B, C, sequence(_, B, C)).
41
42 first(A, sequence(A, _, _)).
43
44 first_name(person(Name, _Car, _Language), Name).
45
46 car(person(_Name, Car, _Language), Car).
47
48 language(person(_Name, _Car, Language), Language).

```

Abbildung 7.6: Wer fährt Toyota?

Die obige Interpretation des gegebenen Prolog-Programms ist *funktional*, dass heißt wir fassen die ersten beiden Argumente des Prädikats `myAppend` als *Eingaben* auf und interpretieren das letzte Argument als *Ausgabe*. Diese Interpretation ist aber keineswegs die einzig mögliche Interpretation. Um das zu sehen, geben wir als Ziel

```
myAppend(L1, L2, [1,2,3]).
```

ein und drücken, nachdem das System uns die erste Antwort gegeben hat, nicht die Taste *Return* sondern die Taste “;”. Wir erhalten:

```
?- myAppend(L1, L2, [1, 2, 3]).
```

```
L1 = []  
L2 = [1, 2, 3] ;
```

```
L1 = [1]  
L2 = [2, 3] ;
```

```
L1 = [1, 2]  
L2 = [3] ;
```

```
L1 = [1, 2, 3]  
L2 = [] ;
```

```
No
```

In diesem Fall hat das Prolog-System durch Backtracking alle Möglichkeiten bestimmt, die es gibt, um die Liste “[1, 2, 3]” in zwei Teillisten zu zerlegen.

7.3.1 Sortieren durch Einfügen

Wir entwickeln nun einen einfachen Algorithmus zum Sortieren von Listen von Zahlen. Die Idee ist Folgende: Um eine Liste aus n Zahlen zu sortieren, sortieren wir zunächst die letzten $n - 1$ Zahlen und fügen dann das erste Element an der richtigen Stelle in die sortierte Liste ein. Mit dieser Idee besteht das Programm aus zwei Prädikaten:

1. Das Prädikat `insert/3` erwartet als erstes Argument eine Zahl x und als zweites Argument eine Liste von Zahlen l , die bereits in aufsteigender Reihenfolge sortiert ist. Das Prädikat fügt die Zahl x so in die Liste l ein, dass die resultierende Liste ebenfalls in aufsteigender Reihenfolge sortiert ist. Das so berechnete Ergebnis wird als letztes Argument des Prädikats `insert/3` zurück gegeben.

Um die obigen Ausführungen über die verwendeten Typen und die Bestimmung von Ein- und Ausgabe prägnanter formulieren zu können, führen wir den Begriff einer *Typ-Spezifikation* ein. Für das Prädikat `insert/3` hat diese Typ-Spezifikation die Form

```
insert(+Number, +List(Number), -List(Number)).
```

Das Zeichen “+” legt dabei fest, dass das entsprechende Argument eine Eingabe ist, während “-” verwendet wird um ein Ausgabe-Argument zu spezifizieren.

2. Das Prädikat `insertion_sort/2` hat die Typ-Spezifikation

```
insertion_sort(+List(Number), -List(Number)).
```

Der Aufruf `insertion_sort(List, Sorted)` sortiert die als Eingabe gegebene Liste *List* in aufsteigender Reihenfolge.

Abbildung 7.7 auf Seite 163 zeigt das *Prolog*-Programm. Nachfolgend diskutieren wir die einzelnen Klauseln der Implementierung des Prädikats `insert`.

```

1  % insert( +Number, +List(Number), -List(Number) ).
2
3  insert( X, [], [ X ] ).
4
5  insert( X, [ Head | Tail ], [ X, Head | Tail ] ) :-
6      X <= Head.
7
8  insert( X, [ Head | Tail ], [ Head | New_Tail ] ) :-
9      X > Head,
10     insert( X, Tail, New_Tail ).
11
12 % insertion_sort( +List(Number), -List(Number) ).
13
14 insertion_sort( [], [] ).
15
16 insertion_sort( [ Head | Tail ], Sorted ) :-
17     insertion_sort( Tail, Sorted_Tail ),
18     insert( Head, Sorted_Tail, Sorted ).

```

Abbildung 7.7: Sortieren durch Einfügen.

1. Die erste Klausel des Prädikats `insert` greift, wenn die Liste, in welche die Zahl `X` eingefügt werden soll, leer ist. In diesem Fall wird als Ergebnis einfach die Liste zurück gegeben, die als einziges Element die Zahl `X` enthält.
2. Die zweite Klausel greift, wenn die Liste, in die die Zahl `X` eingefügt werden soll, nicht leer ist und wenn außerdem `X` kleiner oder gleich dem ersten Element dieser Liste ist. In diesem Fall kann `X` an den Anfang der Liste gestellt werden. Dann erhalten wir die Liste

`[X, Head | Tail]`.

Diese Liste ist sortiert, weil einerseits schon die Liste `[Head | Tail]` sortiert ist und andererseits `X` kleiner als `Head` ist.

3. Die dritte Klausel greift, wenn die Liste, in die die Zahl `X` eingefügt werden soll nicht leer ist und wenn außerdem `X` größer als das erste Element dieser Liste ist. In diesem Fall muss `X` rekursiv in die Liste `Tail` eingefügt werden. Dabei bezeichnet `Tail` den Rest der Liste, in die wir `X` einfügen wollen. Weiter bezeichnet `New_Tail` die Liste, die wir erhalten, wenn wir die Zahl `X` in die Liste `Tail` einfügen. An den Anfang der Liste `New_Tail` setzen wir nun noch den Kopf `Head` der als Eingabe gegebenen Liste.

Damit können wir nun auch die Wirkungsweise des Prädikats `insertion_sort` erklären.

1. Ist die zu sortierende Liste leer, so ist das Ergebnis die leere Liste.
2. Ist die zu sortierende Liste nicht leer und hat die Form `[Head | Tail]`, so sortieren wir zunächst die Liste `Tail` und erhalten als Ergebnis die sortierte Liste `Sorted_Tail`. Fügen wir hier noch das Element `Head` mit Hilfe von `insert` ein, so erhalten wir als Endergebnis die sortierte Liste.

Viele *Prolog*-Prädikate sind *funktional*. Wir nennen ein Prädikat funktional, wenn die einzelnen Argumente klar in Eingabe- und Ausgabe-Argumente unterschieden werden können und wenn außerdem zu jeder Eingabe höchstens eine Ausgabe berechnet wird. Zum Beispiel sind die oben angegebenen Prädikate zum Sortieren einer Liste von Zahlen funktional. Bei einem funktionalen Programm können wir die Semantik oft dadurch am besten verstehen, dass wir das Programm in *bedingte Gleichungen* umformen. Für das oben angegebene Programm erhalten wir dann die folgenden Gleichungen:

1. $\text{insert}(X, []) = [X]$.
2. $X \leq \text{Head} \rightarrow \text{insert}(X, [\text{Head}|\text{Tail}]) = [X, \text{Head}|\text{Tail}]$.
3. $X > \text{Head} \rightarrow \text{insert}(X, [\text{Head}|\text{Tail}]) = [\text{Head}|\text{insert}(X, \text{Tail})]$.
4. $\text{insertion_sort}([]) = []$.
5. $\text{insertion_sort}([\text{Head}|\text{Tail}]) = \text{insert}(\text{Head}, \text{insertion_sort}(\text{Tail}))$.

Die Korrespondenz zwischen dem *Prolog*-Programm und den Gleichungen sollte augenfällig sein. Außerdem ist offensichtlich, dass die obigen Gleichungen den Sortier-Algorithmus in sehr prägnanter Form wiedergeben. Wir werden diese Beobachtung im zweiten Semester benutzen und die meisten dort vorgestellten Algorithmen durch bedingte Gleichungen spezifizieren.

7.3.2 Sortieren durch Mischen

Der im letzten Abschnitt vorgestellte Sortier-Algorithmus hat einen Nachteil: Die Rechenzeit, die dieser Algorithmus verbraucht, wächst im ungünstigsten Fall quadratisch mit der Länge der zu sortierenden Liste. Den Beweis dieser Behauptung werden wir im nächsten Semester liefern. Wir werden nun einen Algorithmus vorstellen der effizienter ist: Ist n die Länge der Liste, so wächst bei diesem Algorithmus der Verbrauch der Rechenzeit nur mit dem Faktor $n \cdot \log_2(n)$. Den Nachweis dieser Behauptung erbringen wir im zweiten Semester. Wenn es sich bei der zu sortierenden Liste beispielsweise um ein Telefonbuch mit einer Millionen Einträgen handelt, dann ist der relative Unterschied zwischen n^2 und $n \log_2(n)$ bei etwa 50 000.

Wir werden den effizienteren Algorithmus zunächst durch bedingte Gleichungen beschreiben und anschließend die Umsetzung dieser Gleichungen in *Prolog* angeben. Der Algorithmus wird in der Literatur als *Sortieren durch Mischen* bezeichnet (engl. *merge sort*) und besteht aus drei Phasen:

1. In der ersten Phase wird die zu sortierende Liste in zwei etwa gleich große Teillisten aufgeteilt.
2. In der zweiten Phase werden diese Teillisten rekursiv sortiert.
3. In der dritten Phase werden die sortierten Teillisten so zusammen gefügt (gemischt), dass die resultierende Liste ebenfalls sortiert ist.

Wir beginnen mit dem Aufteilen einer Liste in zwei Teile. Bei der Aufteilung orientieren wir uns an den Indizes der Elemente. Zur Illustration zunächst ein Beispiel: Wir teilen die Liste

$$[a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8] \quad \text{auf in} \quad [a_1, a_3, a_5, a_7] \quad \text{und} \quad [a_2, a_4, a_6, a_8].$$

Elemente, deren Index gerade ist, werden in der ersten Teilliste aufgesammelt und die Elemente mit ungeradem Index sammeln wir in der zweiten Teilliste. Als Namen für die Funktionen, die diese Teillisten berechnen, wählen wir **even** und **odd**:

$$\begin{aligned} \text{odd} &: \text{List}(\text{Number}) \rightarrow \text{List}(\text{Number}), \\ \text{even} &: \text{List}(\text{Number}) \rightarrow \text{List}(\text{Number}). \end{aligned}$$

Die Funktion **odd**(L) berechnet die Liste aller Elemente aus L mit ungeradem Index, während **even**(L) die Liste aller Elemente mit geradem Index berechnet. Die beiden Funktionen können durch die folgenden Gleichungen spezifiziert werden:

1. $\text{odd}([]) = []$.
2. $\text{odd}([h|t]) = [h|\text{even}(t)]$,
denn das erste Element einer Liste hat den Index 1, was offenbar ein ungerader Index ist und alle Elemente, die in der Liste t einen geraden Index haben, haben in der Liste $[h|t]$ einen ungeraden Index.
3. $\text{even}([]) = []$.

4. $\text{even}([h|t]) = \text{odd}(t)$,

denn alle Elemente, die in der Liste t einen ungeraden Index haben, haben in der Liste $[h|t]$ einen geraden Index.

Als nächstes entwickeln wir eine Funktion

$$\text{mix} : \text{List}(\text{Number}) \times \text{List}(\text{Number}) \rightarrow \text{List}(\text{Number})$$

die zwei aufsteigend sortierte Listen so mischt, dass die resultierende Liste ebenfalls aufsteigend sortiert ist. Durch rekursive Gleichungen kann diese Funktion wie folgt spezifiziert werden:

1. $\text{mix}([], l) = l$.

2. $\text{mix}(l, []) = l$.

3. $x \leq y \rightarrow \text{mix}([x|s], [y|t]) = [x|\text{mix}(s, [y|t])]$.

Falls $x \leq y$ ist, so ist x sicher das kleinste Element der Liste, die entsteht, wenn wir die bereits sortierten Listen $[x|s]$ und $[y|t]$ mischen. Also mischen wir rekursiv die Listen s und $[y|t]$ und setzen x an den Anfang dieser Liste.

4. $x > y \rightarrow \text{mix}([x|s], [y|t]) = [y|\text{mix}([x|s], t)]$.

Damit können wir jetzt die Funktion

$$\text{merge_sort} : \text{List}(\text{Number}) \rightarrow \text{List}(\text{Number}),$$

die eine Liste von Zahlen sortiert, durch bedingte Gleichungen spezifizieren.

1. $\text{merge_sort}([]) = []$.

2. $\text{merge_sort}([x]) = [x]$.

3. $\text{length}(l) \geq 2 \rightarrow \text{merge_sort}(l) = \text{mix}(\text{merge_sort}(\text{odd}(l)), \text{merge_sort}(\text{even}(l)))$.

Falls die Liste l aus 2 oder mehr Elementen besteht, teilen wir diese Liste in die beiden Listen $\text{odd}(l)$ und $\text{even}(l)$ auf, sortieren diese Listen und mischen anschließend die sortierten Teillisten.

Die oben angegebenen Gleichungen lassen sich nun unmittelbar in ein *Prolog*-Programm umsetzen. Abbildung 7.8 auf Seite 166 zeigt das resultierende *Prolog*-Programm. Da es in *SWI-Prolog* bereits vordefinierte Prädikate mit den Namen `merge/3` und `sort/2` gibt, habe ich statt dessen die Namen `mix/2` und `merge_sort/3` gewählt.

7.3.3 Symbolisches Differenzieren

Die Sprache *Prolog* wird gerne für Anwendungen benutzt, bei denen symbolische Rechnungen eine wesentliche Rolle spielen, denn symbolische Rechnungen sind in *Prolog* dadurch, dass die zu manipulierenden Objekte in der Regel unmittelbar als Prolog-Terme dargestellt werden können, sehr einfach zu implementieren. Zur Verdeutlichung zeigen wir ein Programm, mit dem es möglich ist, symbolisch zu differenzieren. Im Rahmen einer Übung haben wir ein SETLX-Programm entwickelt, das arithmetische Ausdrücke symbolisch differenziert. Damals hatten wir davon profitiert, dass die Sprache SETLX Terme als Datenstruktur zur Verfügung stellt. Da in der Sprache *Prolog* Terme ebenfalls fest eingebaut sind, ist es in *Prolog* genauso einfach, symbolisch zu differenzieren.

Die Methodik, mit der wir das *Prolog*-Programm entwickeln, besteht aus zwei Schritten:

1. Als erstes legen wir fest, was genau wir unter einem arithmetischen Ausdruck verstehen wollen und wie ein solcher Ausdruck in *Prolog* repräsentiert werden soll. Dazu definieren wir die Menge der *Prolog*-Terme *Expr*, die einen arithmetischen Ausdruck darstellen.

```

1  % odd( +List(Number), -List(Number) ).
2  odd( [], [] ).
3  odd( [ X | Xs ], [ X | L ] ) :-
4      even( Xs, L ).
5
6  % even( +List(Number), -List(Number) ).
7  even( [], [] ).
8  even( [ _X | Xs ], L ) :-
9      odd( Xs, L ).
10
11 % merge( +List(Number), +List(Number), -List(Number) ).
12 mix( [], Xs, Xs ).
13 mix( Xs, [], Xs ).
14 mix( [ X | Xs ], [ Y | Ys ], [ X | Rest ] ) :-
15     X <= Y,
16     mix( Xs, [ Y | Ys ], Rest ).
17 mix( [ X | Xs ], [ Y | Ys ], [ Y | Rest ] ) :-
18     X > Y,
19     mix( [ X | Xs ], Ys, Rest ).
20
21 % merge_sort( +List(Number), -List(Number) ).
22 merge_sort( [], [] ).
23 merge_sort( [ X ], [ X ] ).
24 merge_sort( [ X, Y | Rest ], Sorted ) :-
25     odd( [ X, Y | Rest ], Odd ),
26     even( [ X, Y | Rest ], Even ),
27     merge_sort( Odd, Odd_Sorted ),
28     merge_sort( Even, Even_Sorted ),
29     mix( Odd_Sorted, Even_Sorted, Sorted ).

```

Abbildung 7.8: Sortieren durch Mischen.

2. Dann stellen wir bedingte Gleichungen auf, die eine Funktion

$$\text{diff} : \text{Expr} \times \text{Var} \rightarrow \text{Expr}$$

beschreiben. Diese Gleichungen sind nichts anderes als die mathematischen Regeln, die Sie in der Schule für das Differenzieren gelernt haben.

3. Im letzten Schritt implementieren wir diese Gleichungen in Prolog.

Induktive Definition der Menge *Expr*.

1. Variablen sind arithmetische Ausdrücke.

Variablen stellen wir durch nullstellige Funktionszeichen dar. Nullstellige Funktionszeichen werden in Prolog auch als *Atome* bezeichnet. Damit gilt

$$c \in \text{Expr} \quad \text{für jedes Prolog-Atom } c.$$

2. Zahlen sind arithmetische Ausdrücke.

Sowohl die ganzen Zahlen als auch die Fließkomma-Zahlen sind Bestandteil der Sprache *Prolog* und können damit durch sich selbst dargestellt werden:

$$n \in Expr \quad \text{für alle } n \in \mathbb{Z},$$

$$r \in Expr \quad \text{für alle } r \in \mathbb{R}.$$

3. Das Negative eines arithmetischen Ausdrucks ist ein arithmetischer Ausdruck. In *Prolog* kann das Negative durch den unären Operator “-” dargestellt werden, also haben wir

$$-t \in Expr \quad \text{falls } t \in Expr.$$

4. Die Summe, die Differenz, das Produkt, und der Quotient zweier arithmetischen Ausdrücke ist ein arithmetischer Ausdruck. In *Prolog* können Summe, Differenz, Produkt und Quotient respektive durch die binären Operatoren “+”, “-”, “*” und “/” dargestellt werden, also setzen wir

$$s + t \in Expr \quad \text{falls } s, t \in Expr.$$

$$s - t \in Expr \quad \text{falls } s, t \in Expr.$$

$$s * t \in Expr \quad \text{falls } s, t \in Expr.$$

$$s / t \in Expr \quad \text{falls } s, t \in Expr.$$

5. Die Potenz zweier arithmetischer Ausdrücke ist ein arithmetischer Ausdruck. In *Prolog* kann die Potenz durch den binären Operator “**” dargestellt werden, also setzen wir

$$s ** t \in Expr \quad \text{falls } s, t \in Expr.$$

6. Bei der Behandlung spezieller Funktionen beschränken wir uns auf die Exponential-Funktion und den natürlichen Logarithmus:

$$\exp(t) \in Expr \quad \text{falls } t \in Expr,$$

$$\ln(t) \in Expr \quad \text{falls } t \in Expr.$$

Aufstellen der bedingten Gleichungen Den Wert von $\text{diff}(t, x)$ definieren wir nun durch Induktion nach dem Aufbau des arithmetischen Ausdrucks t .

1. Bei der Ableitung einer Variablen müssen wir unterscheiden, ob wir die Variable nach sich selbst oder nach einer anderen Variablen ableiten.

- (a) Die Ableitung einer Variablen nach sich selbst gibt den Wert 1:

$$y = x \rightarrow \frac{dy}{dx} = 1.$$

Also haben wir

$$y = x \rightarrow \text{diff}(y, x) = 1.$$

- (b) Die Ableitung einer Variablen y nach einer anderen Variablen x ergibt den Wert 0:

$$y \neq x \rightarrow \frac{dy}{dx} = 0$$

Also haben wir

$$y \neq x \rightarrow \text{diff}(y, x) = 0.$$

2. Die Ableitung einer Zahl n ergibt 0:

$$\frac{dn}{dx} = 0.$$

Damit haben wir

$$\mathbf{diff}(n, x) = 0.$$

3. Die Ableitung eines Ausdrucks mit negativen Vorzeichen ist durch

$$\frac{d}{dx}(-f) = -\frac{df}{dx}$$

gegeben. Die rekursive Gleichung lautet

$$\mathbf{diff}(-f, x) = -\mathbf{diff}(f, x).$$

4. Die Ableitung einer Summe ergibt sich als Summe der Ableitungen der Summanden:

$$\frac{d}{dx}(f + g) = \frac{df}{dx} + \frac{dg}{dx}$$

Als Gleichung schreibt sich dies

$$\mathbf{diff}(f + g, x) = \mathbf{diff}(f, x) + \mathbf{diff}(g, x).$$

5. Die Ableitung einer Differenz ergibt sich als Differenz der Ableitung der Operanden:

$$\frac{d}{dx}(f - g) = \frac{df}{dx} - \frac{dg}{dx}$$

Als Gleichung schreibt sich dies

$$\mathbf{diff}(f - g, x) = \mathbf{diff}(f, x) - \mathbf{diff}(g, x).$$

6. Die Ableitung eines Produktes wird durch die Produkt-Regel beschrieben:

$$\frac{d}{dx}(f * g) = \frac{df}{dx} * g + f * \frac{dg}{dx}.$$

Dies führt auf die Gleichung

$$\mathbf{diff}(f * g, x) = \mathbf{diff}(f, x) * g + f * \mathbf{diff}(g, x).$$

7. Die Ableitung eines Quotienten wird durch die Quotienten-Regel beschrieben:

$$\frac{d}{dx}(f/g) = \frac{\frac{df}{dx} * g - f * \frac{dg}{dx}}{g * g}.$$

Dies führt auf die Gleichung

$$\mathbf{diff}(f/g, x) = (\mathbf{diff}(f, x) * g - f * \mathbf{diff}(g, x)) / (g * g).$$

8. Zur Ableitung eines Ausdrucks der Form $f ** g$ verwenden wir die folgende Gleichung:

$$f ** g = \exp(g * \ln(f)).$$

Das führt auf die Gleichung

$$\text{diff}(f ** g, x) = \text{diff}(\exp(g * \ln(f)), x).$$

9. Bei der Ableitung der Exponential-Funktion benötigen wir die Ketten-Regel:

$$\frac{d}{dx} \exp(f) = \frac{df}{dx} * \exp(f).$$

Das führt auf die Gleichung

$$\text{diff}(\exp(f), x) = \text{diff}(f, x) * \exp(f).$$

10. Für die Ableitung des natürlichen Logarithmus finden wir unter Berücksichtigung der Ketten-Regel

$$\frac{d}{dx} \ln(f) = \frac{1}{f} * \frac{df}{dx}.$$

Das führt auf die Gleichung

$$\text{diff}(\ln(f), x) = \text{diff}(f, x) / f.$$

Implementierung in Prolog Abbildung 7.9 zeigt die Implementierung in *Prolog*. An Stelle der zweistelligen Funktion *diff()* haben wir nun ein dreistelliges Prädikat *diff/3*, dessen letztes Argument das Ergebnis berechnet. Wir diskutieren die einzelnen Klauseln.

1. Die beiden Klauseln in den Zeilen 3 – 9 zeigen, wie eine Variable differenziert werden kann. Das Prädikat *atom(X)* prüft, ob *X* ein nullstelliges Funktions-Zeichen ist. Solche Funktions-Zeichen werden im *Prolog*-Jargon auch als *Atome* bezeichnet. Wir prüfen also in Zeile 4 und 8, ob es sich bei dem abzuleitenden Ausdruck um eine Variable handelt. Anschließend überprüfen wir in den Zeilen 5 bzw. 9, ob diese Variable mit der Variablen, nach der differenziert werden soll, übereinstimmt oder nicht.
2. In der Klausel in den Zeilen 11 – 12 behandeln wir den Fall, dass es sich bei dem zu differenzierenden Ausdruck um eine Zahl handelt. Um dies überprüfen zu können, verwenden wir das Prädikat *number(X)*, das überprüft, ob das Argument *X* eine Zahl ist.

In dieser Klausel haben wir die Variable, nach der abgeleitet werden soll, mit “*X*” bezeichnet. Der Grund ist, dass das *Prolog*-System für Variablen, die in einer Klausel nur einmal vorkommen, eine Warnung ausgibt. Diese Warnung kann vermieden werden, wenn vorne an den Variablennamen ein Unterstrichs “*_*” angefügt wird.

3. Am Beispiel der Ableitung des Ausdrucks $-f$ zeigen wir, wie rekursive Gleichungen in *Prolog* umgesetzt werden können. Die Gleichung, die in den Zeilen 14 – 15 umgesetzt wird, lautet

$$\text{diff}(-f, x) = -\text{diff}(f, x).$$

Um den Ausdruck $-F$ nach x zu differenzieren, müssen wir zunächst den Ausdruck F nach x ableiten. Das passiert in Zeile 15 und liefert das Ergebnis Fs . Das Endergebnis erhalten wir dadurch, dass wir vor Fs ein Minuszeichen setzen.

4. Die restlichen Klausel setzen die oben gefundenen bedingten Gleichungen unmittelbar um und werden daher hier nicht weiter diskutiert.

```

1  % diff( +Expr, +Atom, -Expr).
2
3  diff(F, X, 1) :-
4      atom(F),
5      F == X.
6
7  diff(F, X, 0) :-
8      atom(F),
9      F \== X.
10
11 diff(N, _X, 0) :-
12     number(N).
13
14 diff(-F, X, -Fs) :-
15     diff(F, X, Fs).
16
17 diff(F + G, X, Fs + Gs) :-
18     diff(F, X, Fs),
19     diff(G, X, Gs).
20
21 diff(F - G, X, Fs - Gs) :-
22     diff(F, X, Fs),
23     diff(G, X, Gs).
24
25 diff(F * G, X, Fs * G + F * Gs) :-
26     diff(F, X, Fs),
27     diff(G, X, Gs).
28
29 diff(F / G, X, (Fs * G - F * Gs) / (G * G)) :-
30     diff(F, X, Fs),
31     diff(G, X, Gs).
32
33 diff( F ** G, X, D ) :-
34     diff( exp(G * ln(F)), X, D ).
35
36 diff( exp(F), X, Fs * exp(F) ) :-
37     diff(F, X, Fs).
38
39 diff( ln(F), X, Fs / F ) :-
40     diff(F, X, Fs).

```

Abbildung 7.9: Ein Programm zum symbolischen Differenzieren

7.4 Negation in *Prolog*

In diesem Abschnitt besprechen wir die Implementierung des Negations-Operators in *Prolog*. Wir zeigen zunächst an Hand eines einfachen Beispiels die Verwendung dieses Operators, besprechen dann seine Semantik und zeigen abschließend, in welchen Fällen die Verwendung des Negations-Operators problematisch ist.

7.4.1 Berechnung der Differenz zweier Listen

In *Prolog* wird der Negations-Operator als “\+” geschrieben. Wir erläutern die Verwendung dieses Operators am Beispiel einer Funktion, die die Differenz zweier Mengen berechnen soll, wobei die Mengen durch Listen dargestellt werden. Wir werden die Funktion

difference : $List(Number) \times List(Number) \rightarrow List(Number)$

durch bedingte Gleichungen spezifizieren. Der Ausdruck

difference(l_1, l_2)

berechnet die Liste aller der Elemente aus l_1 , die nicht Elemente der Liste l_2 sind. In SETLX könnten wir diese Funktion wie in Abbildung 7.10 gezeigt implementieren.

```
1  difference := procedure(l1, l2) {
2      return [ x in l1 | !(x in l2) ];
3  };
```

Abbildung 7.10: Implementierung der Prozedur **difference** in SETLX.

In *Prolog* erfolgt die Implementierung dieser Funktion durch Rekursion im ersten Argument. Dazu stellen wir zunächst bedingte Gleichungen auf:

1. $difference([], l) = []$.
2. $\neg member(h, l) \rightarrow difference([h|t], l) = [h|difference(t, l)]$,
denn wenn das Element h in der Liste l nicht vorkommt, so bleibt dieses Element im Ergebnis erhalten.
3. $member(h, l) \rightarrow difference([h|t], l) = difference(t, l)$.

```
1  % difference( +List(Number), +List(Number), -List(Number) ).
2  difference( [], _L, [] ).
3
4  difference( [ H | T ], L, [ H | R ] ) :-
5      \+ member( H, L ),
6      difference( T, L, R ).
7
8  difference( [ H | T ], L, R ) :-
9      member( H, L ),
10     difference( T, L, R ).
```

Abbildung 7.11: Berechnung der Differenz zweier Listen

7.4.2 Semantik des Negations-Operators in Prolog

Es bleibt zu klären, wie das *Prolog*-System eine Anfrage der Form

`\+ A`

beantwortet, wie also der **not**-Operator implementiert ist.

1. Zunächst versucht das System, die Anfrage “A” zu beantworten.
2. Falls die Beantwortung der Anfrage “A” scheitert, ist die Beantwortung der Anfrage “\+ A” erfolgreich. In diesem Fall werden keine Variablen instanziiert.
3. Falls die Beantwortung der Anfrage “A” erfolgreich ist, so scheitert die Beantwortung der Anfrage “\+ A”.

Wichtig ist zu sehen, dass bei der Beantwortung einer negierten Anfrage in keinem Fall Variablen instanziiert werden. Eine negierte Anfrage

`\+ A`

funktioniert daher nur dann wie erwartet, wenn die Anfrage *A* keine Variablen mehr enthält. Zur Illustration betrachten wir das Programm in Abbildung 7.12. Versuchen wir mit diesem Programm die Anfrage

`smart1(X)`

zu beantworten, so wird diese Anfrage reduziert zu der Anfrage

`\+ roemer(X), gallier(X).`

Um die Anfrage “\+ roemer(X)” zu beantworten, versucht das *Prolog*-System rekursiv, die Anfrage “roemer(X)” zu beantworten. Dies gelingt und die Variable *X* wird dabei an den Wert “caesar” gebunden. Da die Beantwortung der Anfrage “roemer(X)” gelingt, scheitert die Anfrage

`\+ roemer(X)`

und damit gibt es auch auf die ursprüngliche Anfrage “smart1(X)” keine Antwort.

```
1  gallier(miraculix).
2
3  roemer(caesar).
4
5  smart1(X) :- \+ roemer(X), gallier(X).
6
7  smart2(X) :- gallier(X), \+ roemer(X).
```

Abbildung 7.12: Probleme mit der Negation

Wenn wir voraussetzen, dass das Programm das Prädikate **roemer**/1 vollständig beschreibt, dann ist dieses Verhalten rein logisch betrachtet nicht korrekt, denn die Konjunktion

`¬roemer(miraculix) ∧ gallier(miraculix)`

folgt aus den im Programm gegebenen Fakten. Wenn der dem *Prolog*-System zu Grunde liegende automatische Beweiser anders implementiert wäre, dann könnte er dies auch erkennen. Wir können uns in diesem Beispiel damit behelfen, dass wir die Reihenfolge der Formeln im Rumpf umdrehen, so wie dies bei der Klausel in Zeile 7 der Abbildung 7.12 geschehen ist. Die Anfrage

`smart2(X)`

liefert für *X* den Wert “miraculix”. Die zweite Anfrage funktioniert, weil zu dem Zeitpunkt, an dem die negierte Anfrage “\+ roemer(X)” aufgerufen wird, ist die Variable *X* bereits an den Wert *miraculix* gebunden und die Anfrage “roemer(miraculix)” scheitert. Generell sollte in *Prolog*-Programmen der Negations-Operator “\+” nur auf solche Prädikate angewendet werden, die zum Zeitpunkt des Aufrufs keine freien Variablen mehr enthalten.

7.4.3 Extralogische Prädikate

Es gibt bestimmte Prädikate, die nicht nach dem an früherer Stelle beschriebenen Algorithmus ausgewertet werden, weil sie nicht als Fakten und Regeln gespeichert sind. Hier handelt es sich um die sogenannten *vordefinierten* Prädikate. Ein einfaches Beispiel ist das Prädikat `writeln/1`, das sein Argument gefolgt von einem Zeilenumbruch ausgibt. Ein interessanteres Beispiel ist das Prädikat `is/2`: Dieses Prädikat dient der Auswertung arithmetischer Ausdrücke. Dabei ist das zweite Argument ein arithmetischer Ausdruck, der ausgewertet werden soll. Dieses Ergebnis wird dann an die Variable, die als erstes Argument übergeben wird, gebunden. Beispielsweise liefert die Anfrage

```
?- is(X, 2 + 3).
```

das Ergebnis

```
X = 5.
```

Das Prädikat `is/2` kann auch als Infix-Operator verwendet werden. Beispielsweise hätten wir die obige Anfrage auch als

```
X is 2 + 3.
```

schreiben können. Es ist wichtig zu wissen, dass der arithmetische Ausdruck, der dem Prädikat `is/2` zur Auswertung übergeben wird, keine ungebundenen Variablen mehr enthalten darf. Beispielsweise liefert die Anfrage

```
?- 3 is 2 + X.
```

nicht etwa das Ergebnis `X = 1` sondern statt dessen die Fehlermeldung

```
ERROR: is/2: Arguments are not sufficiently instantiated.
```

Es wäre schön, wenn *Prolog* auch einfach Anfragen wie die obere korrekt beantworten könnte und die obige Gleichung nach `X` auflösen könnte. Es gibt tatsächlich Erweiterungen von *Prolog*, die dazu in der Lage sind. Diese Disziplin wird als *Constrained-Logic-Programming* bezeichnet. Neben dem Prädikat `is/2` haben auch die Prädikate zum Größenvergleich zweier Werte, also die Prädikate `>/2`, `</2`, `>=/2`, `<=/2` die Einschränkung, dass die beteiligten Argumente vollständig instanziiert sein müssen.

7.5 Die Tiefen-Suche in *Prolog*

Wenn das *Prolog*-System eine Anfrage beantwortet, wird dabei als Such-Strategie die sogenannte Tiefen-Suche (engl. *depth first search*) angewendet. Wir wollen diese Strategie nun an einem weiteren Beispiel verdeutlichen. Wir implementieren dazu ein *Prolog*-Programm mit dessen Hilfe es möglich ist, in einem Graphen eine Verbindung von einem gegebenen Start-Knoten zu einem Ziel-Knoten zu finden. Als Beispiel betrachten wir den Graphen in Abbildung 7.13. Die Kanten können durch ein *Prolog*-Prädikat `edge/2` wie folgt dargestellt werden:

```
1  edge(a, b).
2  edge(a, c).
3  edge(b, e).
4  edge(e, f).
5  edge(c, f).
```

Wir wollen nun ein *Prolog*-Programm entwickeln, mit dem es möglich ist, für zwei vorgegebene Knoten x und y zu entscheiden, ob es einen Weg von x nach y gibt. Außerdem soll dieser Weg dann als Liste von Knoten berechnet werden. Unser erster Ansatz besteht aus dem Programm, das in Abbildung 7.14 gezeigt ist. Die Idee ist, dass der Aufruf

```
find_path(Start, Goal, Path)
```

einen Pfad *Path* berechnet, der von *Start* nach *Goal* führt. Wir diskutieren die Implementierung.

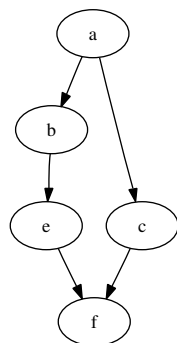


Abbildung 7.13: Ein einfacher Graph ohne Zykeln

```

1  % find_path( +Point, +Point, -List(Point) ).
2  find_path( X, X, [ X ] ).
3
4  find_path( X, Z, [ X | Path ] ) :-
5      edge( X, Y ),
6      find_path( Y, Z, Path ).

```

Abbildung 7.14: Berechnung von Pfaden in einem Graphen

1. Die erste Klausel sagt aus, dass es trivialerweise einen Pfad von X nach X gibt. Dieser Pfad enthält genau den Knoten X .
2. Die zweite Klausel sagt aus, dass es einen Weg von X nach Z gibt, wenn es zunächst eine direkte Verbindung von X zu einem Knoten Y gibt und wenn es dann von diesem Knoten Y eine Verbindung zu dem Knoten

Z gibt. Wir erhalten den Pfad, der von **X** nach **Z** führt, dadurch, dass wir vorne an den Pfad, der von **Y** nach **Z** führt, den Knoten **X** anfügen.

Stellen wir an das *Prolog*-System die Anfrage `find_path(a,f,P)`, so erhalten wir die Antwort

```
1  ?- find_path(a,f,P).
2
3  P = [a, b, e, f] ;
4  P = [a, c, f] ;
5  No
```

Durch Backtracking werden also alle möglichen Wege von **a** nach **b** gefunden. Als nächstes testen wir das Programm mit dem in Abbildung 7.15 gezeigten Graphen. Diesen Graphen stellen wir wie folgt in *Prolog* dar:

```
1  edge(a, b).
2  edge(a, c).
3  edge(b, e).
4  edge(e, a).
5  edge(e, f).
6  edge(c, f).
```

Jetzt erhalten wir auf die Anfrage `find_path(a,f,P)` die Antwort

```
1  ?- find_path(a,f,P).
2  ERROR: Out of local stack
```

Die Ursache ist schnell gefunden.

1. Wir starten mit der Anfrage
`find_path(a,f,P).`
2. Nach Unifikation mit der zweiten Klausel haben wir die Anfrage reduziert auf
`edge(a, Y1), find_path(Y1, f, P1).`
3. Nach Unifikation mit dem Fakt `edge(a,b)` haben wir die neue Anfrage
`find_path(b, f, P1).`
4. Nach Unifikation mit der zweiten Klausel haben wir die Anfrage reduziert auf
`edge(b, Y2), find_path(Y2, f, P2).`
5. Nach Unifikation mit dem Fakt `edge(b,e)` haben wir die neue Anfrage
`find_path(e, f, P2).`
6. Nach Unifikation mit der zweiten Klausel haben wir die Anfrage reduziert auf
`edge(e, Y3), find_path(Y3, f, P3).`
7. Nach Unifikation mit dem Fakt `edge(e,a)` haben wir die neue Anfrage
`find_path(a, f, P3).`

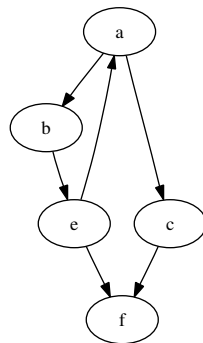


Abbildung 7.15: Ein Graph mit einem Zykel

Die Anfrage “`find_path(a, f, P3)`” unterscheidet sich von der ursprünglichen Anfrage “`find_path(a, f, P)`” nur durch den Namen der Variablen. Wenn wir jetzt weiterrechnen würden, würde sich die Rechnung nur wiederholen, ohne dass wir vorwärts kommen. Das Problem ist, das *Prolog* immer die erste Klausel nimmt, die passt. Wenn später die Reduktion der Anfrage scheitert, wird zwar nach Backtracking die nächste Klausel ausprobiert, aber wenn das Programm in eine Endlos-Schleife läuft, dann gibt es eben kein Backtracking, denn das Programm weiß ja nicht, dass es in einer Endlos-Schleife ist.

Es ist leicht das Programm so umzuschreiben, dass keine Endlos-Schleife mehr auftreten kann. Die Idee ist, dass wir uns merken, welche Knoten wir bereits besucht haben und diese nicht mehr auswählen. In diesem Sinne implementieren wir nun ein Prädikat `find_path/4`. Die Idee ist, dass der Aufruf

```
find_path(Start, Goal, Visited, Path)
```

einen Pfad berechnet, der von *Start* nach *Goal* führt und der zusätzlich keine Knoten benutzt, die bereits in der Liste *Visited* aufgeführt sind. Diese Liste füllen wir bei den rekursiven Aufrufen nach und nach mit den Knoten an, die wir bereits besucht haben. Mit Hilfe dieser Liste vermeiden wir es, einen Knoten zweimal zu besuchen. [Abbildung 7.16](#) zeigt die Implementierung.

```

1  % find_path( +Point, +Point, +List(Point), -List(Point) )
2
3  find_path( X, X, _Visited, [ X ] ).
4
5  find_path( X, Z, Visited, [ X | Path ] ) :-
6      edge( X, Y ),
7      \+ member( Y, Visited ),
8      find_path( Y, Z, [ Y | Visited ], Path ).

```

Abbildung 7.16: Berechnung von Pfaden in zyklischen Graphen

1. In der ersten Klausel spielt das zusätzliche Argument noch keine Rolle, denn wenn wir das Ziel erreicht haben, ist es uns egal, welche Knoten wir schon besucht haben.
2. In der zweiten Klausel überprüfen wir in Zeile 7, ob der Knoten *Y* in der Liste *Visited*, die die Knoten enthält, die bereits besucht wurden, auftritt. Nur wenn dies nicht der Fall ist, versuchen wir rekursiv von *Y* einen Pfad nach *Z* zu finden. Bei dem rekursiven Aufruf erweitern wir die Liste *Visited* um den Knoten *Y*, denn diesen Knoten wollen wir in Zukunft ebenfalls vermeiden.

Mit dieser Implementierung ist es jetzt möglich, auch in dem zweiten Graphen einen Weg von **a** nach **f** zu finden, wir erhalten folgendes Ergebnis:

```

1  ?- find_path(a,f,[a],P).
2  P = [a, b, e, f] ;
3  P = [a, c, f] ;
4  No

```

7.5.1 Die Bekehrung der Ungläubigen

Als spielerische Anwendung zeigen wir nun, wie sich mit Hilfe des oben definierten Prädikats `find_path/4` ein theologisches Problem lösen läßt.

Drei Missionare und drei Ungläubige wollen zusammen einen Fluß überqueren. Sie haben nur ein Boot, indem maximal zwei Passagiere fahren können. Sowohl die Ungläubigen als auch die Missionare können rudern. Weder die Gläubigen, noch die Ungläubigen können über das Wasser laufen. Die Ungläubigen sind hungrig, wenn die Missionare an einem der Ufer in der Unterzahl sind, haben sie ein Problem. Die Aufgabe besteht darin, einen Fahrplan zu erstellen, so dass hinterher alle das andere Ufer erreichen und die Missionare zwischendurch kein Problem haben.

Die Idee ist, das Rätsel, durch einen Graphen zu modellieren. Die Knoten dieses Graphen sind dann die Situationen, die während des Übersetzens auftreten. Wir repräsentieren diese Situationen durch Terme der Form

`side(M, K, B)`.

Ein solcher Term repräsentiert eine Situation, bei der auf der linken Seite des Ufers M Missionare, K Ungläubige und B Boote sind. Unsere Aufgabe besteht nun darin, das Prädikat `edge/2` so zu implementieren, dass

`edge(side(M1, K1, B1), side(M2, K2, B2))`

genau dann wahr ist, wenn die Situation `side(M1, K1, B1)` durch eine Boots-Überfahrt in die Situation `side(M2, K2, B2)` überführt werden kann und wenn zusätzlich die Missionare in der neuen Situation kein Problem bekommen. Abbildung 7.18 auf Seite 180 zeigt ein *Prolog*-Programm, was das Rätsel löst. Den von diesem Programm berechneten Fahrplan finden Sie in Abbildung 7.17 auf Seite 179. Wir diskutieren dieses Programm nun Zeile für Zeile.

1. Wir beginnen mit dem Hilfs-Prädikat `otherSide/4`, das in den Zeilen 24 – 26 implementiert ist. Für eine vorgegebene Situation `side(M, K, B)` berechnet der Aufruf

`otherSide(side(M, K, B), OtherSide)`

einen Term, der die Situation am gegenüberliegenden Ufer beschreibt. Wenn an einen Ufer M Missionare sind, so sind am anderen Ufer die restlichen Missionare und da es insgesamt 3 Missionare gibt, sind das $3 - M$. Die Anzahl der Ungläubige am gegenüberliegenden Ufer wird analog berechnet.

2. Das Prädikat `problem/2` in den Zeilen 29 – 34 überprüft, ob es bei einer vorgegeben Anzahl von Missionaren und Ungläubige zu einem Problem kommt. Da das Problem entweder am linken oder am rechten Ufer auftreten kann, besteht die Implementierung aus zwei Klauseln. Die erste Klausel prüft, ob es auf der Seite, an der M Missionare und K Ungläubige sind, zum Problem kommt. Die zweite Klausel überprüft, ob es auf dem gegenüberliegenden Ufer zu einem Problem kommt. Als Hilfs-Prädikat verwenden wir hier das Prädikat `problemSide/2`. Dieses Prädikat ist in Zeile 37 implementiert und überprüft die Situation an einer Seite: Falls sich auf einer Seite M Missionare und K Ungläubige befinden, so gibt es dann ein Problem, wenn die Zahl M von 0 verschieden ist und wenn zusätzlich $M < K$ ist.

3. Bei der Implementierung des Prädikats `edge/2` verwenden wir in den Zeilen 9 und 10 das Prädikat `between/3`, das in dem *SWI-Prolog*-System vordefiniert ist. Beim Aufruf

`between(Low, High, N)`

sind Low und $High$ ganze Zahlen mit $Low \leq High$. Der Aufruf instantiiert die Variable N nacheinander mit den Zahlen

1	MMM	KKK	B	~~~~~			
2				> KK >			
3	MMM	K		~~~~~	B	KK	
4				< K <			
5	MMM	KK	B	~~~~~		K	
6				> KK >			
7	MMM			~~~~~	B	KKK	
8				< K <			
9	MMM	K	B	~~~~~		KK	
10				> MM >			
11	M	K		~~~~~	B	KK	MM
12				< M K <			
13	MM	KK	B	~~~~~		K	M
14				> MM >			
15		KK		~~~~~	B	K	MMM
16				< K <			
17		KKK	B	~~~~~			MMM
18				> KK >			
19		K		~~~~~	B	KK	MMM
20				< K <			
21		KK	B	~~~~~		K	MMM
22				> KK >			
23				~~~~~	B	KKK	MMM

Abbildung 7.17: Fahrplan für Missionare und Ungläubige

Low, Low + 1, Low + 2, ..., High.

Beispielsweise gibt die Anfrage

```
between(1,3,N), write(N), nl, fail.
```

nacheinander die Zahlen 1, 2 und 3 am Bildschirm aus.

- Die Implementierung des Prädikats `edge/2` besteht aus zwei Klauseln. In der ersten Klausel betrachten wir den Fall, dass das Boot am linken Ufer ist. In der Zeilen 9 generieren wir die Zahl der Missionare `MB`, die im Boot übersetzen sollen. Diese Zahl `MB` ist durch `M` beschränkt, denn es können nur die Missionare übersetzen, die sich am linken Ufer befinden. Daher benutzen wir das Prädikat `between/3` um eine Zahl zwischen 0 und `M` zu erzeugen. Analog generieren wir in Zeile 10 die Zahl `KB` der Ungläubige, die im Boot übersetzen. In Zeile 11 testen wir, dass es mindestens einen Passagier gibt, der mit dem Boot übersetzt und in Zeile 12 testen wir, dass es höchstens zwei Passagiere sind. In Zeile 13 und 14 berechnen wir die Zahl `MN` der Missionare und die Zahl `KN` der Ungläubige, die nach der Überfahrt auf dem linken Ufer verbleiben und testen dann in Zeile 15, dass es für diese Zahlen kein Problem gibt.

Die zweite Klausel befaßt sich mit dem Fall, dass das Boot am rechten Ufer liegt. Wir hätten diese Klausel mit *Copy & Paste* aus der vorhergehenden Klausel erzeugen können, aber es ist eleganter, diesen Fall auf den vorhergehenden Fall zurück zu führen. Da da Boot nun auf der rechten Seite liegt, berechnen wir daher in Zeile 19 die Zahl `MR` der Missionare auf der rechten Seite und die Zahl `KR` der Ungläubige auf der rechten Seite. Dann untersuchen wir die Situation `side(MR,KR,1)`, bei der `MR` Missionare und `KR` Ungläubige am linken Ufer stehen. Wenn diese so übersetzen können, dass nachher `MRN` Missionare und `KRN` Ungläubige am linken Ufer stehen, dann können wir in Zeile 21 berechnen, wieviele Missionare und Ungläubige sich dann am gegenüberliegenden Ufer befinden.

- In den Zeilen 1 – 4 definieren wir nun das Prädikat `solve/0`, dessen Aufruf das Problem löst. Dazu wird zunächst das Prädikat `find_path/4` mit dem Start-Knoten `side(3,3,1)` und dem Ziel-Knoten

```

1  solve :-
2      find_path( side(3,3,1), side(0,0,0), [ side(3,3,1) ], Path ),
3      nl, write('Lösung:' ), nl, nl,
4      print_path(Path).
5
6  % edge( +Point, -Point ).
7  % This clause describes rowing from the left side to the right side.
8  edge( side( M, K, 1 ), side( MN, KN, 0 ) ) :-
9      between( 0, M, MB ),      % MB missionaries in the boat
10     between( 0, K, KB ),      % KB infidels in the boat
11     MB + KB >= 1,             % boat must not be empty
12     MB + KB <= 2,             % no more than two passengers
13     MN is M - MB,             % missionaries left on the left side
14     KN is K - KB,             % infidels left on the left side
15     \+ problem( MN, KN ).     % no problem must occur
16
17 % This clause describes rowing from the right side to the left side.
18 edge( side( M, K, 0 ), side( MN, KN, 1 ) ) :-
19     otherSide( M, K, MR, KR ),
20     edge( side( MR, KR, 1 ), side( MRN, KRN, 0 ) ),
21     otherSide( MRN, KRN, MN, KN ).
22
23 % otherSide( +Number, +Number, -Number, -Number ).
24 otherSide( M, K, M_Other, K_Other ) :-
25     M_Other is 3 - M,
26     K_Other is 3 - K.
27
28 % problem( +Number, +Number ).
29 problem(M, K) :-
30     problemSide(M, K).
31
32 problem(M, K) :-
33     otherSide( M, K, M_Other, K_Other ),
34     problemSide(M_Other, K_Other).
35
36 % problemSide( +Number, +Number ).
37 problemSide(Missionare, Kannibalen) :-
38     Missionare > 0,
39     Missionare < Kannibalen.
40
41 % find_path( +Point, +Point, +List(Point), -List(Point) )
42 find_path( X, X, _Visited, [ X ] ).
43
44 find_path( X, Z, Visited, [ X | Path ] ) :-
45     edge( X, Y ),
46     \+ member( Y, Visited ),
47     find_path( Y, Z, [ Y | Visited ], Path ).

```

Abbildung 7.18: Die Bekehrung der Ungläubigen

`side(0,0,0)` aufgerufen. Der berechnete Pfad wird dann ausgegeben mit dem Prädikat `print_path/1`, dessen Implementierung hier aus Platzgründen nicht angegeben wird.

7.6 Der Cut-Operator

Wir haben ein Prädikat als *funktional* definiert, wenn wir die einzelnen Argumente klar in Eingabe- und Ausgabe-Argumente aufteilen können. Wir nennen ein Prädikat *deterministisch* wenn es funktional ist und wenn außerdem zu jeder Eingabe höchstens eine Ausgabe berechnet wird. Diese zweite Forderung ist durchaus nicht immer erfüllt. Betrachten wir die ersten beiden Fakten zur Definition des Prädikats `mix/3`:

```

1  mix( [], Xs, Xs ).
2  mix( Xs, [], Xs ).
```

Für die Anfrage “`mix([], [], L)`” können beide Fakten verwendet werden. Das Ergebnis ist zwar immer das selbe, nämlich `L = []`, es wird aber zweimal ausgegeben:

```

1  ?- mix([], [], L).
2
3  L = [] ;
4
5  L = []
```

Dies kann zu Ineffizienz führen. Aus diesem Grunde gibt es in *Prolog* den Cut-Operator “!”. Mit diesem Operator ist es möglich, redundante Lösungen aus dem Suchraum heraus zu schneiden. Schreiben wir die ersten beiden Klauseln der Implementierung von `mix/3` in der Form

```

1  mix( [], Xs, Xs ) :- !.
2  mix( Xs, [], Xs ) :- !.
```

so wird auf die Anfrage “`mix([], [], L)`” die Lösung `L = []` nur noch einmal generiert. Ist allgemein eine Regel der Form

$$P :- Q_1, \dots, Q_m, !, R_1, \dots, R_k$$

gegeben, und gibt es weiter eine Anfrage A , so dass A und P unifizierbar sind, so wird die Anfrage A zunächst zu der Anfrage

$$Q_1\mu, \dots, Q_m\mu, !, R_1\mu, \dots, R_k\mu$$

reduziert. Außerdem wird ein Auswahl-Punkt gesetzt, wenn es noch weitere Klauseln gibt, deren Kopf mit A unifiziert werden könnte. Bei der weiteren Abarbeitung dieser Anfrage gilt folgendes:

1. Falls bereits die Abarbeitung einer Anfrage der Form

$$Q_i\sigma, \dots, Q_m\sigma, !, R_1\sigma, \dots, R_k\sigma$$

für ein $i \in \{1, \dots, m\}$ scheitert, so wird der Cut nicht erreicht und hat keine Wirkung.

2. Eine Anfrage der Form

$$!, R_1\sigma, \dots, R_k\sigma$$

wird reduziert zu

$$R_1\sigma, \dots, R_k\sigma.$$

Dabei werden alle Auswahl-Punkte, die bei der Beantwortung der Teilanfragen Q_1, \dots, Q_m gesetzt worden sind, gelöscht. Außerdem wird ein eventuell bei der Reduktion der Anfrage A auf die Anfrage

$$Q_1\mu, \dots, Q_m\mu, !, R_1\mu, \dots, R_k\mu$$

gesetzter Auswahl-Punkte gelöscht.

3. Sollte später die Beantwortung der Anfrage

$$R_1\sigma, \dots, R_k\sigma.$$

scheitern, so scheitert auch die Beantwortung der Anfrage A .

Zur Veranschaulichung betrachten wir ein Beispiel.

```

1  q(Z) :- p(Z).
2  q(1).
3
4  p(X) :- a(X), b(X), !, c(X,Y), d(Y).
5  p(3).
6
7  a(1).    a(2).    a(3).
8
9  b(2).    b(3).
10
11 c(2,2).  c(2,4).
12
13 d(3).

```

Wir verfolgen die Beantwortung der Anfrage $q(U)$.

1. Zunächst wird versucht $q(U)$ mit dem Kopf der ersten Klausel des Prädikats $q/1$ zu unifizieren. Dabei wird Z mit U instantiiert und die Anfrage wird reduziert zu

$$p(U).$$

Da es noch eine weitere Klausel für das Prädikat $q/1$ gibt, die zur Beantwortung der Anfrage $q(U)$ in Frage kommt, setzen wir Auswahl-Punkt Nr. 1.

2. Jetzt wird versucht $p(U)$ mit $p(X)$ zu unifizieren. Dabei wird die Variable X an U gebunden und die ursprüngliche Anfrage wird reduziert zu der Anfrage

$$a(U), b(U), !, c(U,Y), d(Y).$$

Außerdem wird an dieser Stelle Auswahl-Punkt Nr. 2 gesetzt, denn die zweite Klausel des Prädikats $p/1$ kann ja ebenfalls mit der ursprünglichen Anfrage unifiziert werden.

3. Um die Teilanfrage $a(U)$ zu beantworten, wird $a(U)$ mit $a(1)$ unifiziert. Dabei wird U mit 1 instantiiert und die Anfrage wird reduziert zu

$$b(1), !, c(1,Y), d(Y).$$

Da es für das Prädikat $a/1$ noch weitere Klauseln gibt, wird Auswahl-Punkt Nr. 3 gesetzt.

4. Jetzt wird versucht, die Anfrage

$$b(1), !, c(1,Y), d(Y).$$

zu lösen. Dieser Versuch scheitert jedoch, da sich die für das Prädikat $b/1$ vorliegenden Fakten nicht mit $b(1)$ unifizieren lassen.

5. Also springen wir zurück zum letzten Auswahl-Punkt (das ist Auswahl-Punkt Nr. 3) und machen die Instantiierung $U \mapsto 1$ rückgängig. Wir haben jetzt also wieder das Ziel

$$a(U), b(U), !, c(U,Y), d(Y).$$

6. Diesmal wählen wir das Fakt $a(2)$ um es mit $a(U)$ zu unifizieren. Dabei wird U mit 2 instantiiert und wir haben die Anfrage

$$b(2), !, c(2,Y), d(Y).$$

Da es noch eine weitere Klausel für das Prädikat $a/1$ gibt, setzen wir Auswahl-Punkt Nr. 4 an dieser Stelle.

7. Jetzt unifizieren wir die Teilanfrage $b(2)$ mit der ersten Klausel für das Prädikat $b/1$. Die verbleibende Anfrage ist

$$!, c(2, Y), d(Y).$$

8. Diese Anfrage wird reduziert zu

$$c(2, Y), d(Y).$$

Außerdem werden bei diesem Schritt die Auswahl-Punkte Nr. 2 und Nr. 4 gelöscht.

9. Um diese Anfrage zu beantworten, unifizieren wir $c(2, Y)$ mit dem Kopf der ersten Klausel für das Prädikat $c/2$, also mit $c(2, 2)$. Dabei erhalten wir die Instantiierung $Y \mapsto 2$. Die Anfrage ist damit reduziert zu

$$d(2).$$

Außerdem setzen wir an dieser Stelle Auswahl-Punkt Nr. 5, denn das Prädikat $c/2$ hat ja noch eine weitere Klausel, die in Frage kommt.

10. Die Anfrage “ $d(2)$ ” scheitert. Also springen wir zurück zum Auswahl-Punkt Nr. 5 und machen die Instantiierung $Y \mapsto 2$ rückgängig. Wir haben also wieder die Anfrage

$$c(2, Y), d(Y).$$

11. Zur Beantwortung dieser Anfrage nehmen wir nun die zweite Klausel der Implementierung von $c/2$ und erhalten die Instantiierung $Y \mapsto 4$. Die verbleibende Anfrage lautet dann

$$d(4).$$

12. Da sich $d(4)$ und $d(3)$ nicht unifizieren lassen, scheitert diese Anfrage. Wir springen jetzt zurück zum Auswahl-Punkt Nr. 1 und machen die Instantiierung $U \mapsto Z$ rückgängig. Die Anfrage lautet also wieder

$$p(U).$$

13. Wählen wir nun die zweite Klausel der Implementierung von $q/1$, so müssen wir $q(U)$ und $q(1)$ unifizieren. Diese Unifikation ist erfolgreich und wir erhalten die Instantiierung $U \mapsto 1$, die die Anfrage beantwortet.

7.6.1 Verbesserung der Effizienz von *Prolog*-Programmen durch den Cut-Operator

In der Praxis wird der Cut-Operator eingesetzt, um überflüssige Auswahl-Punkte zu entfernen und dadurch die Effizienz eines Programms zu steigern. Als Beispiel betrachten wir eine Implementierung des Algorithmus “*Sortieren durch Vertauschen*” (engl. *bubble sort*). Wir spezifizieren diesen Algorithmus zunächst durch bedingte Gleichungen. Dabei benutzen wir die Funktion

$$\text{append} : \text{List}(\text{Number}) \times \text{List}(\text{Number}) \rightarrow \text{List}(\text{Number})$$

Der Aufruf $\text{append}(l_1, l_2)$ liefert eine Liste, die aus allen Elementen von l_1 gefolgt von den Elementen aus l_2 besteht. In dem *SWI-Prolog*-System ist ein entsprechendes Prädikat $\text{append}/3$ implementiert. Die Implementierung dieses Prädikats deckt sich mit der Implementierung des Prädikats $\text{concat}/3$, die wir in einem früheren Abschnitt vorgestellt hatten.

Außerdem benutzen wir noch das Prädikat

$$\text{ordered} : \text{List}(\text{Number}) \rightarrow \mathbb{B},$$

das überprüft, ob eine Liste geordnet ist. Die bedingten Gleichungen zur Spezifikation der Funktion

$$\text{bubble_sort} : \text{List}(\text{Number}) \rightarrow \text{List}(\text{Number})$$

lauten nun:

1. $\text{append}(l_1, [x, y|l_2], l) \wedge x > y \rightarrow \text{bubble_sort}(l) = \text{bubble_sort}(\text{append}(l_1, [y, x|l_2]))$
Wenn die Liste l in zwei Teile l_1 und $[x, y|l_2]$ zerlegt werden kann und wenn weiter $x > y$ ist, dann vertauschen wir die Elemente x und y und sortieren die so entstandene Liste rekursiv.
2. $\text{ordered}(l) \rightarrow \text{bubble_sort}(l) = l$
Wenn die Liste l bereits sortiert ist, dann kann die Funktion bubble_sort diese Liste unverändert als Ergebnis zurück geben.

Die Gleichungen um das Prädikat `ordered` zu spezifizieren lauten:

1. `ordered([]) = true`

Die leere Liste ist offensichtlich sortiert.

2. `ordered([x]) = true`

Eine Liste, die nur aus einem Element besteht, ist ebenfalls sortiert.

3. $x \leq y \rightarrow \text{ordered}([x, y|r]) = \text{ordered}([y|r])$.

Eine Liste der Form $[x, y|r]$ ist sortiert, wenn $x \leq y$ ist und wenn außerdem die Liste $[y|r]$ sortiert ist.

```

1  bubble_sort( L, Sorted ) :-
2      append( L1, [ X, Y | L2 ], L ),
3      X > Y,
4      append( L1, [ Y, X | L2 ], Cs ),
5      bubble_sort( Cs, Sorted ).
6
7  bubble_sort( Sorted, Sorted ) :-
8      is_ordered( Sorted ).
9
10
11 is_ordered( [] ).
12
13 is_ordered( [ _ ] ).
14
15 is_ordered( [ X, Y | Ys ] ) :-
16     X < Y,
17     is_ordered( [ Y | Ys ] ).

```

Abbildung 7.19: Der Bubble-Sort Algorithmus

Abbildung 7.19 zeigt die Implementierung des Bubble-Sort Algorithmus in *Prolog*. In Zeile 2 wird die als Eingabe gegebene Liste `L` in die beiden Liste `L1` und `[X, Y | L2]` zerlegt. Da es im Allgemeinen mehrere Möglichkeiten gibt, eine Liste in zwei Teillisten zu zerlegen, wird hierbei ein Auswahl-Punkt gesetzt. Anschließend wird geprüft, ob `Y` kleiner als `X` ist. Wenn dies der Fall ist, wird mit `append/3` die neue Liste

`append(L1, [Y, X|L2])`

gebildet und diese Liste wird rekursiv sortiert. Wenn es nicht möglich ist, die Liste `L` so in zwei Listen `L1` und `[X, Y | L2]` zu zerlegen, dass `Y` kleiner als `X` ist, dann muss die Liste `L` schon sortiert sein. In diesem Fall greift die zweite Klausel, die allerdings noch überprüfen muss, ob `L` tatsächlich sortiert ist, denn sonst könnte beim Backtracking eine falsche Lösung berechnet werden.

Das Problem bei dem obigen Programm ist die Effizienz. Aufgrund der vielen Möglichkeiten eine Liste zu zerlegen, wird beim Backtracking immer wieder die selbe Lösung generiert. Beispielsweise liefert die Anfrage

`bubble_sort([4, 3, 2, 1], L), write(L), nl, fail.`

16 mal die selbe Lösung. Abbildung 7.20 zeigt eine Implementierung, bei der nur eine Lösung berechnet wird. Dies wird durch den Cut-Operator in Zeile 4 erreicht. Ist einmal eine Zerlegung der Liste `L` der Form

$L = L_1 + [X, Y|L_2]$

gefunden, bei der `Y` kleiner als `X` ist, so bringt es nichts mehr, nach anderen Zerlegungen zu suchen, denn die ursprünglich gegebene Liste `L` lässt sich ja auf jeden Fall dadurch sortieren, dass rekursiv die Liste

$L_1 + [Y, X|L_2]$

sortiert wird. Dann kann auch der Aufruf der Prädikats `ordered/1` im Rumpf der zweiten Klausel des Prädikats `/2bubble_sort` entfallen, denn diese wird beim Backtracking ja nur dann erreicht, wenn es keine Zerlegung der Liste `L` in `L1` und `[X, Y | L2]` gibt, bei der `Y` kleiner als `X` ist. Dann muß aber die Liste `L` schon sortiert sein.

```

1  bubble_sort( List, Sorted ) :-
2      append( L1, [ X, Y | L2 ], List ),
3      X > Y,
4      !,
5      append( L1, [ Y, X | L2 ], Cs ),
6      bubble_sort( Cs, Sorted ).
7
8  bubble_sort( Sorted, Sorted ).

```

Abbildung 7.20: Effiziente Implementierung des Bubble-Sort Algorithmus

Wenn wir bei der Entwicklung eines *Prolog*-Programms von bedingten Gleichungen ausgehen, dann gibt es ein einfaches Verfahren, um das entstandene *Prolog*-Programm durch die Einführung von Cut-Operator effizienter zu machen: Der Cut-Operator sollte nach den Tests, die vor dem Junktor “ \rightarrow ” stehen, gesetzt werden. Wir erläutern dies durch ein Beispiel: Unten sind noch einmal die Gleichungen zur Spezifikation des Algorithmus “*Sortieren durch Mischen*” wiedergegeben.

1. `odd([]) = []`.
2. `odd([h|t]) = [h|even(t)]`.
3. `even([]) = []`.
4. `even([h|t]) = odd(t)`.
5. `merge([], l) = l`.
6. `merge(l, []) = l`.
7. $x \leq y \rightarrow \text{merge}([x|s], [y|t]) = [x|\text{merge}(s, [y|t])]$.
8. $x > y \rightarrow \text{merge}([x|s], [y|t]) = [y|\text{merge}([x|s], t)]$.
9. `sort([]) = []`.
10. `sort([x]) = [x]`.
11. `sort([x, y|t]) = merge(sort(odd([x, y|t])), sort(even([x, y|t])))`.

Das *Prolog*-Programm mit Cut-Operatoren sieht dann so aus wie in Abbildung 7.21 gezeigt. Nur die Gleichungen 7. und 8. haben Bedingungen, bei allen anderen Gleichungen gibt es keine Bedingungen. Bei den Gleichungen 7. und 8. wird der Cut-Operator daher nach dem Test der Bedingungen gesetzt, bei allen anderen Klauseln wird der Cut-Operator dann am Anfang des Rumpfes gesetzt.

Analysieren wir das obige Programm genauer, so stellen wir fest, dass viele der Cut-Operatoren im Grunde überflüssig sind. Beispielsweise kann immer nur eine der beiden Klauseln, die das Prädikat `odd/2` implementieren, greifen, denn entweder ist die eingegebene Liste leer oder nicht. Also sind die Cut-Operatoren in den Zeilen 2 und 4 redundant. Andererseits stören sie auch nicht, so dass es für die Praxis das einfachste sein dürfte Cut-Operatoren stur nach dem oben angegebenen Rezept zu setzen.

```

1  % odd( +List(Number), -List(Number) ).
2  odd( [], [] ) :- !.
3  odd( [ X | Xs ], [ X | L ] ) :-
4      !,
5      even( Xs, L ).
6
7  % even( +List(Number), -List(Number) ).
8  even( [], [] ) :- !.
9  even( [ _X | Xs ], L ) :-
10     !,
11     odd( Xs, L ).
12
13 % merge( +List(Number), +List(Number), -List(Number) ).
14 mix( [], Xs, Xs ) :- !.
15 mix( Xs, [], Xs ) :- !.
16 mix( [ X | Xs ], [ Y | Ys ], [ X | Rest ] ) :-
17     X <= Y,
18     !,
19     mix( Xs, [ Y | Ys ], Rest ).
20 mix( [ X | Xs ], [ Y | Ys ], [ Y | Rest ] ) :-
21     X > Y,
22     !,
23     mix( [ X | Xs ], Ys, Rest ).
24
25 % merge_sort( +List(Number), -List(Number) ).
26 merge_sort( [], [] ) :- !.
27 merge_sort( [ X ], [ X ] ) :- !.
28 merge_sort( [ X, Y | Rest ], Sorted ) :-
29     !,
30     odd( [ X, Y | Rest ], Odd ),
31     even( [ X, Y | Rest ], Even ),
32     merge_sort( Odd, Odd_Sorted ),
33     merge_sort( Even, Even_Sorted ),
34     mix( Odd_Sorted, Even_Sorted, Sorted ).

```

Abbildung 7.21: Sortieren durch Mischen mit Cut-Operatoren.

7.7 Literaturhinweise

Für eine umfangreiche und dem Thema angemessene Darstellung der Sprache *Prolog* fehlt in der einführenden Vorlesung leider die Zeit. Daher wird dieses Thema in einer späteren Vorlesung auch wieder aufgegriffen. Den Lesern, die ihre Kenntnisse jetzt schon vertiefen wollen, möchte ich die folgende Hinweise auf die Literatur geben:

1. *The Art of Prolog* von Leon Sterling und Ehud Shapiro [SS94]. Dieses Werk ist ein ausgezeichnete Lehrbuch, das auch für den Anfänger gut lesbar ist.
2. *Prolog Programming for Artificial Intelligence* von Ivan Bratko [Bra90]. Neben der Sprache *Prolog* führt dieses Buch auch in die künstliche Intelligenz ein.
3. *Foundations of Logic Programmming* von J. W. Lloyd [Llo87] beschreibt die theoretischen Grundlagen der Sprache *Prolog*.
4. *Prolog: The Standard* von Pierre Deransart, Abdel Ali Ed-Dbali und Laurent Cervoni [DEDC96] gibt den ISO-Standard für die Sprache *Prolog* wieder.
5. *SWI-Prolog 6.2.6 Reference Manual* von Jan Wielemaker [Wie13] beschreibt das SWI-Prolog-System. Dieses Dokument ist im Internet unter der Adresse

<http://www.swi-prolog.org/download/stable/doc/SWI-Prolog-6.2.6.pdf>

verfügbar.

Literaturverzeichnis

- [Bra90] Ivan Bratko. *PROLOG Programming for Artificial Intelligence*. Addison-Wesley, 2nd edition, 1990.
- [Can95] Georg Cantor. Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen*, 46:481–512, 1895.
- [DEDC96] P. Deransart, A. Ed-Dbali, and L. Cervoni. *Prolog: The Standard*. Springer-Verlag, 1996.
- [DLL62] Martin Davis, George Logemann, and Donald Loveland. A machine program for theorem-proving. *Communications of the ACM*, 5(7):394–397, 1962.
- [Lip98] Seymour Lipschutz. *Set Theory and Related Topics*. McGraw-Hill, New York, 1998.
- [Llo87] J. W. Lloyd. *Foundations of Logic Programming*. Springer, Berlin, 1987.
- [McC97] William McCune. Solution of the robbins problem. *Journal of Automated Reasoning*, 19:263–276, December 1997.
- [McC10] William McCune. Prover9 and Mace4. <http://www.cs.unm.edu/~mccune/prover9/>, 2005–2010.
- [MMZ⁺01] Matthew W. Moskewicz, Conor F. Madigan, Ying Zhao, Lintao Zhang, and Sharad Malik. Chaff: Engineering an Efficient SAT Solver. In *Proceedings of the 38th Design Automation Conference (DAC'01)*, 2001.
- [Ric53] Henry G. Rice. Classes of recursively enumerable sets and their decision problems. *Transactions of the American Mathematical Society*, 83, 1953.
- [Sch87] Uwe Schöning. *Logik für Informatiker*. Springer-Verlag, 1987.
- [SDSD86] J. T. Schwartz, R. B. Dewar, E. Schonberg, and E. Dubinsky. *Programming With Sets: An Introduction to SETL*. Springer-Verlag, 1986.
- [SS94] Leon Sterling and Ehud Y. Shapiro. *The Art of Prolog - Advanced Programming Techniques*, 2nd Ed. MIT Press, 1994.
- [Tur36] Alan M. Turing. On computable numbers, with an application to the “Entscheidungsproblem”. *Proceedings of the London Mathematical Society*, 42(2):230–265, 1936.
- [Wie13] Jan Wielemaker. SWI Prolog Reference Manual, *updated for version 6.2.6*. Technical report, University of Amsterdam, 2013.