Skriptum zur Vorlesung

Einführung in die Theoretische Informatik

Georg Moser

Wintersemester 2021/22

 10. Auflage

Inhaltsverzeichnis

Teil I.

Einleitung & Beweismethoden

Einleitung

In der Vorlesung „Einführung in die Theoretische Informatik“ werden die folgenden Themen behandelt—in alphabetischer Reihenfolge: (i) *Aussagenlogik*; (ii) *Berechenbarkeitstheorie*; (iii) *Boolesche Algebra*; (iv) *Chomsky-Hierarchie*; (v) *Formale Sprachen*; (vi) *Komplexitätstheore*; (vii) *Programmverifikation*; (viii) *Reduktionen*; (ix) *Registermaschinen*; (x) *Turingmaschinen.* Diese Themen werden in der SL „Einführung in die Theoretische Informatik“ in weiterführenden Übungen vertieft.

Das vorliegende Skriptum struktiert diese Themen in zwei Teile und stellt diesen noch eine Einleitung und eine Zusammenfassung von oft verwendeten Beweismethoden voran, sowie einen Anhang nach, sodass das Skriptum insgesamt der folgenden Struktur folgt:

- Teil I: Einleitung & Beweismethoden
- Teil II: Logik, Algebra & Formale Sprachen
- Teil III: Berechenbarkeit, Komplexität & Verifikation
- Teil IV: Anhang

Der erste Teil besteht hauptsächlich aus dem Kapitel „Beweismethoden“. In diesem Kapitel werden gängige Beweismethoden eingeführt, welche wir im Laufe der Vorlesung verwenden werden. Dieses Kapitel dient zur Vertiefung des Verständnisses der verwendeten Beweisprinzipien und ist nicht Teil der Vorlesung. Das Skriptum setzt ein minimales Verständnis formaler Konzepte voraus, wie sie etwa in einer allgemein bildenden höheren Schule vermittelt, beziehungsweise im Brückenkurs Mathematik wiederholt werden.

Im Teil II beginnt der Hauptteil des Skriptums. Hier werden die Themen (i) *Aussagenlogik*; (ii) *Boolesche Algebra*; (iii) *Chomsky-Hierarchie* und (iv) *Formale Sprachen*. behandelt. Der Teil „Logik, Algebra & Formale Sprachen“ vermittelt gewissermaßen die Basiswerkzeuge der theoretischen Informatikerin. Im Kapitel 2 wird die *Aussagenlogik* besprochen. Im Kapitel 3 werden *Algebren* im Allgemeinen sowie die *Boolesche Algebra* besprochen. Im Kapitel 4 werden *Grammatiken*, die *Chomsky-Hierarchie* und Grundlagen der *formalen Sprachen* eingeführt.

Im Teil III „Berechenbarkeit, Komplexität & Verifikation“, bauen wir teilweise auf diesen Werkzeugen auf und skizzieren bedeutsame Probleme und Techniken aus den folgenden Bereichen: (i) *Berechenbarkeitstheorie*; (ii) *Komplexitätstheore* und (iii) *Programmverifikation*. Im Kapitel 5 werden abstrakte Berechnunsmodelle dargestellt und die Theorie der *Berechenbarkeit* skizziert. Im Kapitel 6 wird kurz auf die Grundbegriffe der *Komplexitätstheorie* eingegangen. Abschließend wird *Programmverifikation* im Kapitel 7 besprochen.

Schließlich nimmt der Anhang Themen auf, die den Zusammenhang zu anderen Disziplinen bzw. Lehrveranstaltungen vertiefen. Dieser Anhang soll interessierten Studierenden Anleitungen zum Selbststudium geben und ist nicht Teil des Prüfungsstoffes. In Kapitel A.1 wird der Zusammenhang zwischen Boolescher Algebra zur Theorie der *Schaltkreise* dargestellt. Im Kapitel A.2 wird eine Anwednung von kontextfreien Grammatiken in der *Wissensrepräsentation* dargestellt.

Das vorliegende Skriptum ersetzt nicht den Besuch der Vorlesung, sondern ist vorlesungsbegleitend konzipiert. Beachten Sie zum Beispiel, dass das Skriptum nur relativ wenige erläuternde Beispiele enthält, diese werden mehrheitlich nur in der Lehrveranstaltung präsentiert. Zur weiteren Übung schließt jedes Kapitel mit einer Aufgabensammlung ab. Ähnliche Aufgaben werden in der, die Vorlesung begleitenden, SL besprochen. Beachten Sie auch, dass im OLAT Kurs der Lehrveranstaltung weitere Selbsttests zum Eigenstudium zur Verfügung stehen.

Die zehnte Auflage des Skriptums stellt eine gründliche Überarbeitung und Simplifzierung des Skriptums dar. Das Ziel war es die bestehenden Inhalte besser an die Vorlesung und andere Lehrveranstaltungen der Fakultät anzupassen. Dazu wurde in der Aussagenlogik der Teil zum formalen Beweisen überarbeitet und die in früheren Auflagen enthaltene universelle Algebra gestrichen. Außerdem habe ich Korrekturen und Umformulierungen vorgenommen. Ich möchte mich herzlich bei Martin Avanzini, Christian Dalvit, Jamie Hochrainer, Johannes Niederhauser und Jonas Schöpf für die Unterstützung bei der Erstellung der zehnten Auflage bedanken!

Um die Lesbarkeit zu erleichtern, wird in der direkten Anrede des Lesers, der Leserin prinzipiell die weibliche Form gewählt. In der Erstellung des Skriptums habe ich mich auf die folgende Literatur gestützt (in der Reihenfolge der Wichtigkeit für dieses Skriptum) [9, 7, 6, 2, 5, 10].

1.

Beweismethoden

1.1. Motivation

Wozu exakte Definitionen?

Die Begrifflichkeiten in der Mathematik und der theoretischen Informatik müssen exakt definiert sein. Dies führt unter anderem zu einer Reduktion auf das Wesentliche und vermeidet Redundanzen. Etwa sind alle Begriffe der *diskreten Strukturen* [7] aus den Begriffen „Menge“ und „Abbildung“ abgeleitet, z.B.

> Nummerierung, Ordnung, Graph, Wort .

Dem Nachteil des Aufwandes für die exakte Definition stehen folgende Vorteile gegenüber:

- Abstrakte Repräsentation
- Gleichheit mit ja oder nein entscheidbar
- Programmierung naheliegend

Wozu Beweise?

- Mit einem ausformulierten Beweis kann man sich selbst oder Kollegen überzeugen, dass richtig überlegt wurde. Oft stellt sich erst im Beweis an Hand der Argumentationskette heraus, dass gewisse Voraussetzungen fehlen oder überflüssig sind. Nach Kurt Gödel können wir sagen, dass “beweisen” einfach “richtig denken” bedeutet.
- Durch das Studium von Beweisen trainieren Sie das logische Denken und werden befähigt, eigene Ideen korrekt zu formulieren und durch einen Beweis zu verifizieren.
- Beweise führen oft zu programmierbaren Verfahren, wenn diese konstruktiv geführt werden, das heißt, wenn aus den Beweisen Algorithmen ablesbar sind.
- In sicherheitskritischen Anwendungen (autonomes Fahren, Medizin, etc.) kann fehlerbehaftete Software Menschen gefährden (und tut dies leider auch). Es ist unabdingbar, bestimmte Eigenschaften von Programmen zu beweisen. Dies wird auch im Kapitel 7 vertieft werden.

1.2. Deduktive Beweise

Ein *deduktiver Beweis* besteht aus einer Folge von Aussagen, die von einer *Hypothese* zu einer *Konklusion* führen. Jeder Beweisschritt muss sich nach einer akzeptierten logischen Regel aus den gegebenen Fakten oder aus vorangegangenen Aussagen ergeben. In Definition 2.6 in Kapitel 2 lernen wir formale, deduktive Beweise kennen.

Der Aussage, dass die Folge der Beweisschritte von einer Hypothese H zu einer Konklusion K führt, entspricht der Satz:

> Wenn H, dann K.

Wir betrachten einen Satz dieser Form.

Satz 1.1

Wenn $x \geqslant 4$, dann $2^x \geqslant x^2$.

Beweisskizze. Für $x = 4$ richtig: $2^4 \geqslant 4^2$. Für $x \geqslant 1$ gilt, dass sich die linke Seite verdoppelt, wenn x um 1 erhöht wird. Die rechte wächst hingegen nur mit $\frac{(x+1)^2}{x^2}$. Gilt nun $x \geqslant 4$, dann muss gelten $\frac{x+1}{x} \leqslant 1{,}25$, und somit $\frac{(x+1)^2}{x^2} \leqslant 1{,}5625 < 2$. □

Die gegebene Argumentation ist akkurat, jedoch informell. Um den Satz (oder besser das Sätzchen) formal beweisen zu können, benötigen wir vollständige Induktion. Wir geben einen Induktionsbeweis in Sektion 1.4. Der folgende Satz folgt mittels einer stringenten Folge von Aussagen aus seiner Hypothese und Satz 1.1.

Satz 1.2

Wenn x die Summe der Quadrate von 4 positiven ganzen Zahlen ist, dann $2^x \geqslant x^2$.

Beweis. Wir listen die notwendige Folge von Aussagen tabellarisch auf.

$x = a^2 + b^2 + c^2 + d^2$	Hypothese	(1.1)
$a \geqslant 1, b \geqslant 1, c \geqslant 1, d \geqslant 1$	Hypothese	(1.2)
$a^2 \geqslant 1, b^2 \geqslant 1, c^2 \geqslant 1, d^2 \geqslant 1$	(1.2) und elementare Arithmetik	(1.3)
$x \geqslant 4$	folgt aus (1.1) und (1.3)	(1.4)
$2^x \geqslant x^2$	(1.4) und voriger Satz 1.1	(1.5)

□

Formen von „Wenn-dann“

„Wenn-dann“-Sätze können auch in anderen Formen auftreten. Beispiele hierzu sind:

- H impliziert K.
- H nur dann, wenn K.
- K, wenn H.

– Wenn H gilt, folgt daraus K.

– $H \to K$.

„Genau dann, wenn“-Sätze

Gelegentlich finden wir Aussagen der Form „A genau dann, wenn B“. Andere Varianten dieses Satzes sind etwa:

– A dann–und nur dann–wenn B.

– $A \approx B$, $A \leftrightarrow B$, $A \equiv B$.

„Genau dann, wenn“ Aussagen werden bewiesen indem *zwei* Behauptungen gezeigt werden:

– A impliziert B und

– B impliziert A.

Beachten Sie bitte den Sprachgebrauch. Der Aussage S

> A *genau dann, wenn* B,

entsprechen die beiden Aussagen „A *nur dann, wenn* B“ (also „A impliziert B“) und „A, *wenn* B“ („B impliziert A“). Abkürzend schreibt man manchmal für die Richtung der Aussage S von links nach rechts „$\Rightarrow$“ und für die Richtung von rechts nach links „$\Leftarrow$“.

1.3. Beweisformen

Bis jetzt haben wir vor allem die Struktur von Sätzen beziehungsweise Beweisen betrachtet. Im Folgenden gehen wir auf häufig bzw. sehr häufig auftretende Beweisprinzipien ein.

Reduktion auf Definitionen

Viele Aussagen folgen leicht oder sogar unmittelbar, sobald die in den Hypothesen verwendeten Begriffe in ihre Definitionen umgewandelt werden. Wir geben dazu ein einfaches Beispiel, das gleichzeitig wichtige Grundbegriffe der elementaren Mengenlehre einführt. Der folgende Beweis liefert auch das erste Beispiel eines Widerspruchsbeweises, ein oftmals sehr nützliches Beweisprinzip.

Satz 1.3

Sei S eine endliche Teilmenge einer unendlichen Menge U und T sei die Komplementärmenge von S in Bezug auf U. Dann ist T unendlich.

Beweis. Laut Definition gilt $S \cup T = U$ und S, T disjunkt, also $|S| + |T| = |U|$, wobei $|S|$ die *Kardinalität* oder *Mächtigkeit* der Menge S angibt.

Da S endlich, existiert eine bestimmte natürliche Zahl n, sodass $|S| = n$. Andererseits, da U unendlich, existiert *kein* l, sodass $|U| = l$. Nun angenommen T ist endlich, dann existiert wiederum eine natürlich Zahl m, sodass $|T| = m$. Daraus folgt, dass $|U| = |S| + |T| = n + m$. Somit würde eine natürliche Zahl $l = n + m$ existieren, sodass die Kardinalität von U gleich l. Das steht jedoch im Widerspruch zur Annahme, dass U unendlich ist. Somit muss die Annahme, dass T endlich ist, falsch sein. Es folgt die Aussage des Satzes. □

Beweis in Bezug auf Mengen

Wir geben das folgende einfache Beispiel.

Satz 1.4: Distributivgesetz der Vereinigung

$$R \cup (S \cap T) = (R \cup S) \cap (R \cup T) \,.$$

Dieser Satz stellt de-facto einen „Genau dann, wenn“-Satz dar.

Beweisansatz. Wir beschreiben hier nur den Beweisansatz. Zunächst formulieren wir die Gleichung zu einem „Genau dann, wenn“-Satz um:

$$x \in R \cup (S \cap T) \text{ gdw } x \in (R \cup S) \cap (R \cup T).$$

Um diesen Satz zeigen zu können müssen wir nun nur noch die folgenden beiden Behauptungen zeigen.

1. $x \in R \cup (S \cap T)$ impliziert $x \in (R \cup S) \cap (R \cup T)$ *und*
2. $x \in (R \cup S) \cap (R \cup T)$ impliziert $x \in R \cup (S \cap T)$

Betrachten wir die folgende Implikation:

$$x \in (R \cup S) \cap (R \cup T) \quad \text{impliziert} \quad x \in R \cup (S \cap T) \,. \tag{1.6}$$

Oft ist es vorteilhaft, statt einer zu beweisenden Implikation, wie der gerade angegebenen, ihre *Kontraposition* zu betrachten. Die Kontraposition von (1.6) ist dann:

$$x \notin R \cup (S \cap T) \quad \text{impliziert} \quad x \notin (R \cup S) \cap (R \cup T) \,.$$

□

Im Beweisansatz haben wir einen typischen Fall aufgelistet. Statt zu zeigen, dass die Aussage A die Aussage B impliziert, wird gezeigt, dass die Negation von B, die Negation von A impliziert. Gerade bei „Genau dann, wenn“ kann diese Umformulierung nützlich sein. Wir stellen den Sachverhalt schematisch dar:

$$\frac{\text{A impliziert B}}{\text{(nicht B) impliziert (nicht A)}} \qquad \frac{\text{(nicht B) impliziert (nicht A)}}{\text{A impliziert B}}$$

Zur Darstellung von Schlussfolgerungen werden wie hier oft *Schlussfiguren*, oder *Inferenzregeln*, verwendet, wobei die Annahme(n) oberhalb und die Konklusion unterhalb einer horizontalen Linie dargestellt werden. Zusammen bedeuten diese Schemata also, dass die Aussagen „A impliziert B“ und „nicht B impliziert nicht A“ äquivalent sind, das heißt aus dem einen Satz folgt der andere und umgekehrt. Wir werden solche Schlussformeln noch intensiv im Kapitel 2 studieren.

Widerspruchsbeweise

Widerspruchsbeweise sind von entscheidender Bedeutung; da wir bereits einen solchen Beweis angegeben haben, begnügen wir uns im folgenden mit der schematischen Darstellung des Beweises. Im Schema schreiben wir die Wahrheitswertkonstante False für den Widerspruch.

Beweisansatz Widerspruchsbeweis.

$$\frac{\text{Hypothese(n)} \quad \text{Negation der Konklusion}}{\frac{\mathsf{False}}{\text{Konklusion}}}$$

□

Gegenbeispiele

Da Sätze *allgemeine* Aussagen behandeln, genügt es, die Aussage für bestimmte Werte zu widerlegen, um den ganzen Satz zu widerlegen. In dieser Situation haben wir dann ein *Gegenbeispiel* gefunden. Gegenbeispiele können auch verwendet werden, um allgemein gefasste Aussagen soweit zu präzisieren, dass sie dann als Satz gezeigt werden können.

Beispiel

Wir betrachten die folgende (falsche) Behauptung, die eine Verschärfung der Aussage des Satzes 1.1 darstellt:

Für alle $x \geqslant 0$: $2^x \geqslant x^2$.

Die Behauptung ist richtig für $x \geqslant 4$, wie oben behauptet und was wir auch in kommenden Abschnitt beweisen werden. Die Behauptung ist allerdings falsch für $x = 3$

$$2^3 = 8 \not\geqslant 9 = 3^2 \ .$$

Somit haben wir ein *Gegenbeispiel* gefunden.

1.4. Induktive Beweise

Induktive Beweise mit ganzen Zahlen

Zunächst behandeln wir Induktionsbeweise mit ganzen Zahlen. Induktionsbeweise sind immer dann erforderlich, wenn eine Aussage $S(n)$ für alle n gezeigt werden soll. In diesem Fall gehen wir wie folgt vor:

- Basis: Zu zeigen, dass S für Startwert gilt, etwa $n = 0$ oder $n = 1$.
- Induktionsschritt: Zu zeigen, dass wenn $S(n)$, dann gilt auch $S(n + 1)$.

Das zugrundeliegende Prinzip wird *Induktionsprinzip* genannt.

> **Induktionsprinzip**: *Wenn wir $S(i)$ bewiesen haben und beweisen können, dass $S(n)$ für alle $n \geqslant i$ $S(n + 1)$ impliziert, dann können wir daraus schließen, dass $S(n)$ für alle $n \geqslant i$ gilt.*

Wir verdeutlichen das Prinzip indem wir Satz 1.1 formal beweisen.

Satz 1.5

Wenn $x \geqslant 4$, dann $2^x \geqslant x^2$.

Beweis.

- Basis: Für $x = 4$ stimmt die Aussage $2^x \geqslant x^2$.
- Schritt: Zu zeigen ist $2^{x+1} \geqslant (x+1)^2$ unter der Voraussetzung $2^x \geqslant x^2$, der Induktionshypothese. Dazu zeigen wir zunächst die folgende Hilfsüberlegung:

$$2x^2 \geqslant (x+1)^2 \,. \tag{1.7}$$

Da $x \geqslant 4$ gilt $x \geqslant 2 + \frac{1}{x}$ und damit auch $x^2 \geqslant 2x + 1$. Somit gilt:

$$2x^2 = x^2 + x^2 \geqslant x^2 + 2x + 1 = (x+1)^2 \,.$$

Somit folgt (1.7). Schließlich zeigen wir $2^{x+1} \geqslant (x+1)^2$ wie folgt:

$$\begin{aligned} 2^{x+1} &= 2 \cdot 2^x \\ &\geqslant 2 \cdot x^2 \\ &\geqslant (x+1)^2 \,. \end{aligned}$$

Hier haben wir in der zweiten Zeile die Induktionshypothese und in der dritten Zeile die Hilfsüberlegung (1.7) angewandt.

□

Wenn wir kurz (und informell) Gebrauch von Quantoren machen, können wir das Prinzip der vollständigen Induktion auch als Schlussfigur anschreiben:

$$\frac{S(i) \quad \forall n \geqslant i \; (S(n) \to S(n+1))}{\forall n \geqslant i \; S(n)} \,.$$

Hier verwenden wir den Allquantor $\forall$ informell als Abkürzung für „für alle“.

Allgemeinere Formen der Induktion

Das oben beschriebene Induktionsprinzip ist, in der angegeben Form, oft nicht ausreichend. Zwei *Erweiterungen* sind besonders nützlich. Zunächst müssen wir uns nicht auf einen Basisfall konzentrieren, sondern können mehrere Basisfälle verwenden. Zum Beispiel

$$S(i), S(i+1), \ldots, S(j) \,.$$

Im Weiteren können wir, um $S(n+1)$ zu beweisen, als Hypothesen alle Aussagen

$$S(i), S(i+1), \ldots, S(n) \,,$$

verwenden. Darüberhinaus, wenn wir mehrere Basisfälle gezeigt haben, dann können wir im Beweis des Induktionsschrittes

$$n \geqslant j \,,$$

annehmen. Zu beachten ist, dass diese *Erweiterungen* des Induktionsprinzips Erweiterungen in der Anwendbarkeit des Prinzips darstellen, aber der Beweisstärke der Induktion über natürlichen Zahlen nichts hinzufügen. Genauer gesagt folgen die oben erwähnten allgemeineren Formen aus der „einfachen“ Induktion.

Induktive Definitionen und Strukturelle Induktion

In der theoretischen Informatik sind wir weniger an Induktionen über natürliche Zahlen interessiert, sondern mehr an Induktionen über den (rekursiv definierten) Strukturen mit denen wir ständig arbeiten, wie etwa Listen, Bäumen oder Ausdrücken. Wir beginnen mit zwei einfachen Beispielen von rekursiv definierten Strukturen, solche Definitionen werden auch als *induktive Definitionen* bezeichnet.

Definition 1.1

Wir definieren *Bäume* induktiv:
Basis: Ein einzelner Knoten ist ein *Baum*; dieser Knoten ist die *Wurzel*.
Schritt: Wenn $T_1, T_2, \ldots, T_k$ Bäume sind, bilden wir einen neuen *Baum* wie folgt:

1. Man beginnt mit einem neuen Knoten N, der die Wurzel des Baumes darstellt.
2. Schließlich fügt man k Kanten von N zu den Wurzeln der T_i hinzu.

Definition 1.2

Ausdrücke:
Basis: Jede Zahl und jeder Buchstabe ist ein *Ausdruck*.
Schritt: Wenn E, F Ausdrücke sind, dann sind auch $E + F$, $E \cdot F$ und (E) *Ausdrücke*.

Die Aussage $S(X)$ soll für alle Strukturen X, die durch eine bestimmte induktive, bzw. rekursive Definition gegeben sind, gezeigt werden. In diesem Fall gehen wir wie folgt vor:

- Basis: Zunächst beweisen wir $S(X)$ für die Basisstruktur(en) X der induktiven Definition.
- Schritt: Wähle Struktur Y, die rekursiv aus $Y_1, Y_2, \ldots, Y_k$ gebildet wird.
  Induktionshyptothese: $S(Y_1), S(Y_2), \ldots, S(Y_k)$ seien wahr.
  Mit Hilfe der Induktionshyptothese wird nun $S(Y)$ gezeigt.

Das zugrundeliegende Induktionsprinzip wird *Strukturelle Induktion* genannt.

> **Strukturelle Induktion**: *Wenn wir $S(X)$ für alle Basisstrukturen X der induktiven Definition beweisen und beweisen können, dass, wenn Y rekursiv mittels $Y_1, Y_2, \ldots, Y_k$ gebildet wurde und $S(Y_1), S(Y_2), \ldots, S(Y_k)$ für die Teilstrukturen $Y_1, Y_2, \ldots, Y_k$ angenommen wird, dann $S(Y)$ folgt, dann können wir daraus schließen dass $S(X)$ für alle nach der induktiven Definition gebildeten Strukturen X gilt.*

Wir zeigen als Beispiel den folgenden Satz mit struktureller Induktion über Bäume.

Satz 1.6

Jeder Baum besitzt genau einen Knoten mehr als Kanten.

Beweis. Die Aussage $S(T)$ lautet: „Wenn T ein Baum ist und n Knoten und e Kanten hat, dann gilt $n = e + 1$."

- Basis: Trivialerweise gilt $n = e + 1$, wenn T nur aus einem Knoten besteht.
- Schritt: Angenommen T habe $T_1, \ldots, T_k$ als direkte Teilbäume und mit Induktionshypothese folgt $S(T_1), \ldots, S(T_k)$.

  Seien nun $n_1, \ldots, n_k$ die Anzahlen der Knoten von $T_1, \ldots, T_k$. Und seien $e_1, \ldots, e_k$ die Anzahlen der Kanten von $T_1, \ldots, T_k$.

  Für alle $i \in [1, k]$ gilt: $n_i = e_i + 1$. Somit

$$n = 1 + n_1 + \cdots + n_k =$$
$$= 1 + (e_1 + 1) + \cdots + (e_k + 1) = k + e_1 + \cdots + e_k + 1 = e + 1 \; .$$

  Somit haben wir auch den Induktionschritt gezeigt und damit den Satz vollständig bewiesen.

□

Teil II.

Logik, Algebra & Formale Sprachen

2.

Einführung in die Logik

In diesem Kapitel wird die *Aussagenlogik* eingeführt. Zum leichteren Verständnis geschieht dies in zwei Stufen. In Abschnitt 2.1 werden gängige Schlussprinzipien, wie sie in allgemeingültigen Argumenten oft vorkommen, besprochen. In Abschnitt 2.2 wird die Sprache der Aussagenlogik formal eingeführt sowie die Semantik der Aussagenlogik definiert. Der Abschnitt 2.3 widmet sich der Bedeutung beziehungsweise der Manipulation von aussagenlogischen Formeln. In Abschnitt 2.4 betrachten wir ein korrektes und vollständiges Beweissystem für die Aussagenlogik. In Abschnitt 2.5 studieren wir Wahrheitsfunktionen und Normalformen von Formeln.

Schließlich gehen wir in Abschnitt 2.6 kurz auf die Bedeutung der Logik für die Informatik ein und skizzieren mögliche Erweiterungen der hier besprochenen Inhalte. Außerdem finden sich in Abschnitt 2.7 (optionale) Aufgaben zu den Themenbereichen dieses Kapitels, die zur weiteren Vertiefung dienen sollen.

2.1. Elementare Schlussweisen

Das Fachgebiet der *Logik* beschäftigt sich ganz allgemein mit der Korrektheit von Argumenten: Wie muss ein Argument aussehen, sodass wir es als allgemeingültig betrachten? Oder, negativ ausgedrückt: Wann ist ein Argument nicht korrekt?

Wie argumentieren wir im täglichen Leben? Betrachten wir beispielsweise die folgende Sequenz von Behauptungen, die die wir wohl als wahr ansehen können:

> *Sokrates ist ein Mensch.*
> *Alle Menschen sind sterblich.*
> *Somit ist Sokrates sterblich.*

Diese Schlussfigur wird *Syllogismus* genannt und wurde bereits in der Antike untersucht. Aus zwei *Prämissen* (auch Ober- und Untersatz genannt), wird ein dritter Satz geschlossen, die *Konklusion*. Im Beispiel ist „Sokrates ist ein Mensch“ der Obersatz, „Alle Menschen sind sterblich“ der Untersatz und schließlich „Somit ist Sokrates sterblich“ die Konklusion. Syllogismen haben immer genau diese Gestalt: Aus zwei Prämissen folgt die Konklusion. Syllogismen können entweder als die Formulierung eines gemeinsamen Satzes verstanden werden, also wenn „Obersatz“ und „Untersatz“, dann „Konklusion“. Äquivalent dazu ist dass wir einen Syllogismus als *Schlussfigur* oder *Inferenz* verstehen: aus „Obersatz“ und „Untersatz“ wird die „Konklusion“ geschlossen.

Das Wichtige bei logischen Schlüssen ist nicht, dass die Prämissen wahr sind, sondern dass die *Schlussfigur* korrekt ist: Wenn die Prämissen wahr sind, dann muss auch die Konklusion wahr sein. Um dies zweifelsfrei behaupten zu können, sucht man nach allgemeinen Strukturen, die Argumentformen aufweisen können.

Schlussfolgerungen wie die in dem oben angegebenen Syllogismus sind ein wenig speziell und verwenden implizit bereits Quantifizierungen über eine Menge von Objekten. Aber ein Argument wie das Folgende, das in der Logik *Modus Ponens* genannt wird, ist intuitiv unmittelbar nachvollziehbar.

> *Wenn das Kind schreit, hat es Hunger.*
> *Das Kind schreit.*
> *Also hat das Kind Hunger.*

Da die Korrektheit des *Modus Ponens* unabhängig vom Wahrheitsgehalt der eigentlichen Aussagen ist, können wir diese Schlussfigur wie folgt verallgemeinern.

> *Wenn A, dann B.*
> *A gilt.*
> *Also gilt B.*

Hierbei stehen A und B für beliebige *Aussagen*, die entweder wahr oder falsch sein können. Weil A und B als Platzhalter für beliebige Aussagen stehen können, sprechen wir auch von *Aussagenvariablen*. Wie in der gerade dargestellten Argumentkette, gilt im Allgemeinen, dass die Korrektheit oder Gültigkeit einer Schlussfigur nicht von den Aussagenvariablen beziehungsweise deren Wahrheitsgehalt abhängt. Nur die Art wie diese Aussagenvariablen verbunden sind, ist wichtig.

Um aus Aussagenvariablen komplexere Sachverhalte aufzubauen, verwendet man sogenannte *Junktoren*. Beispiele für Junktoren wären etwa die *Negation* (symbolisch $\neg$), *Konjunktion* ($\wedge$)

$\neg$	
T	F
F	T

$\wedge$	T	F
T	T	F
F	F	F

$\vee$	T	F
T	T	T
F	T	F

$\rightarrow$	T	F
T	T	F
F	T	T

Abbildung 2.1.: Wahrheitstabellen

und *Disjunktion* ($\vee$) sowie die *Implikation* ($\rightarrow$). Die Bedeutung dieser Symbole wird durch die *Wahrheitstafeln* in Abbildung 2.1 definiert. Hier schreiben wir T für eine wahre Aussage und F für eine falsche.

Mit Hilfe dieser Junktoren lässt sich nun der *Modus Ponens* konzise fassen und mit Hilfe der Wahrheitstafeln (oder Wahrheitstabellen) in Abbildung 2.1 überprüft man leicht die Allgemeingültigkeit dieser Schlussfigur. Üblicherweise schreibt man den *Modus Ponens* als Inferenzregel, wie folgt.

$$\frac{A \rightarrow B \quad A}{B}$$

Diese Schreibweise trennt die beiden Prämissen $A \rightarrow B$ und A durch einen Querstrich von der Konklusion B. Wir können uns diese Regel wie eine Rechenregel vorstellen: Haben wir uns von A und $A \rightarrow B$ überzeugt, dann können wir auch B schließen.

Während uns die Wahrheitstafeln die *Bedeutung* (auch die *Semantik*) der Junktoren angeben, erlaubt eine Inferenzregel die *syntaktische* Manipulation mit Aussagenvariablen oder zusammengesetzten Aussagen. In weiterer Folge definieren wir die Sprache der Aussagenlogik formal und geben einen korrekten und vollständigen Kalkül der Aussagenlogik an.

2.2. Syntax und Semantik der Aussagenlogik

Gegeben sei eine unendliche Menge AT von *atomaren Formeln* (kurz *Atome* genannt), deren Elemente mit p, q, r, ... bezeichnet werden.

Definition 2.1: Syntax der Aussagenlogik

Die *Wahrheitswertsymbole* der Aussagenlogik sind

$$\mathsf{True} \qquad \mathsf{False}\,,$$

und die *Junktoren* der Aussagenlogik sind

$$\neg \qquad \wedge \qquad \vee \qquad \rightarrow\,.$$

Hier ist $\neg$ der einzige unäre Operator und die anderen Operatoren sind alle zweistellig. Die *Formeln* der Aussagenlogik sind induktiv definiert:

1. Eine atomare Formel p ist eine Formel,
2. die Wahrheitswertsymbole (True, False) sind Formeln und,

3. wenn A und B Formeln sind, dann sind

$$\neg A \qquad (A \wedge B) \qquad (A \vee B) \qquad (A \to B)\,,$$

auch Formeln.

Konvention

Zur Erleichterung der Lesbarkeit werden die Klammern um binäre Junktoren oft weggelassen. Dies ist möglich, wenn die folgende Präzedenz der Junktoren gilt: Der Operator $\neg$ bindet stärker als $\vee$ und $\wedge$, welche wiederum stärker als $\to$ binden. Zur Vereinfachung der Darstellung nutzen wir auch, dass die binären Junktoren $\vee$ und $\wedge$ assoziativ und kommutativ sind. Manchmal verwenden wir auch die Konvention, dass $\to$ rechts-assoziativ geklammert wird: $A \to B \to C$ ist gleichbedeutend zu $A \to (B \to C)$. (Wir erinnern uns daran, dass für die Aussagenvariablen A, B, C beliebige Aussagen eingesetzt werden können.)

Die Namen der in Definition 2.1 verwendeten Junktoren wurden schon in Abschnitt 2.1 eingeführt. Die Wahrheitswertsymbole ($\mathsf{True}, \mathsf{False}$) dienen der syntaktischen Repräsentation der Wahrheitswerte T und F. Damit ist die Sprache der Aussagenlogik, also die *Syntax* vollständig definiert.

Wir schreiben T und F für die beiden betrachteten *Wahrheitswerte*. Wie schon in Abschnitt 2.1 bezeichnet T eine wahre und F eine falsche Aussage.

Definition 2.2: Wahrheitswert

Eine *Belegung* $\mathsf{v}\colon \mathsf{AT} \to \{\mathsf{T}, \mathsf{F}\}$ ist eine Abbildung, die Atome mit Wahrheitswerten assoziiert. Wir schreiben $\bar{\mathsf{v}}(A)$ für den *Wahrheitswert* einer Formel A. Der Wahrheitswert $\bar{\mathsf{v}}(A)$ ist induktiv definiert als die Erweiterung der Belegung v mit Hilfe der Wahrheitstafeln in Abbildung 2.1.

$$\bar{\mathsf{v}}(p) = \mathsf{v}(p) \qquad \bar{\mathsf{v}}(\mathsf{True}) = \mathsf{T} \qquad \bar{\mathsf{v}}(\mathsf{False}) = \mathsf{F}$$

$$\bar{\mathsf{v}}(\neg A) = \begin{cases} \mathsf{T} & \bar{\mathsf{v}}(A) = \mathsf{F} \\ \mathsf{F} & \bar{\mathsf{v}}(A) = \mathsf{T} \end{cases}$$

$$\bar{\mathsf{v}}(A \wedge B) = \begin{cases} \mathsf{T} & \bar{\mathsf{v}}(A) = \bar{\mathsf{v}}(B) = \mathsf{T} \\ \mathsf{F} & \text{sonst} \end{cases}$$

$$\bar{\mathsf{v}}(A \vee B) = \begin{cases} \mathsf{F} & \bar{\mathsf{v}}(A) = \bar{\mathsf{v}}(B) = \mathsf{F} \\ \mathsf{T} & \text{sonst} \end{cases}$$

$$\bar{\mathsf{v}}(A \to B) = \begin{cases} \mathsf{T} & \bar{\mathsf{v}}(A) = \mathsf{F} \text{ oder } \bar{\mathsf{v}}(B) = \mathsf{T} \\ \mathsf{F} & \text{sonst} \end{cases}$$

Eine Formel A für die es zumindest eine Belegung gibt, sodass $\bar{\mathsf{v}}(A) = \mathsf{T}$ nennt man *erfüllbar*. Gibt es keine Belegung, sodass $\bar{\mathsf{v}}(A) = \mathsf{T}$, nennt man A *unerfüllbar*.

Definition 2.3: Konsequenzrelation

Die *Konsequenzrelation* $\{A_1, \ldots, A_n\} \models B$ (oder kurz *Konsequenz*), beschreibt, dass $\bar{\mathsf{v}}(B) = \mathsf{T}$, wenn gilt $\bar{\mathsf{v}}(A_1) = \mathsf{T}, \ldots, \bar{\mathsf{v}}(A_n) = \mathsf{T}$ für jede Belegung v. Wir schreiben $\models A$, statt $\varnothing \models A$. Gilt $\models A$, dann heißt die Formel A eine *Tautologie*, beziehungsweise *gültig*. Außerdem schreiben wir der Einfachheit halber oft $A_1, \ldots, A_n \models B$, statt $\{A_1, \ldots, A_n\} \models B$.

Informell bezeugt die Konsequenzrelation, dass aus der Wahrheit der Prämissen A_1, …, A_n, die Wahrheit der Konklusion B folgt.

Satz 2.1

Eine Formel A ist eine Tautologie gdw.[a] $\neg A$ unerfüllbar ist.

[a] Wir schreiben gdw. als Abkürzung für „genau dann, wenn“.

Beweis. Wir zerlegen den „genau, dann wenn“-Satz in zwei Implikationen, die wir getrennt zeigen, siehe auch die Erklärungen zu Beweismethoden im Anhang.

1. Wir zeigen zunächst die Richtung von links nach rechts. Angenommen A ist eine Tautologie, dann gilt $\bar{\mathsf{v}}(A) = \mathsf{T}$ für alle Belegungen v, also im Besonderen gilt $\bar{\mathsf{v}}(\neg A) = \mathsf{F}$ für alle Belegungen v. Somit ist $\neg A$ unerfüllbar.

2. Nun zeigen wir die Richtung von rechts nach links. Angenommen $\neg A$ ist unerfüllbar. Dann gilt für alle Belegungen v, dass $\bar{\mathsf{v}}(\neg A) = \mathsf{F}$. Somit gilt für alle Belegungen, dass $\bar{\mathsf{v}}(A) = \mathsf{T}$ und A ist eine Tautologie.

□

Um die Gültigkeit einer Formel A beziehungsweise ihre Erfüllbarkeit oder Unerfüllbarkeit festzustellen, genügt es, alle möglichen Belegungen v zu betrachten und die entsprechenden Wahrheitswerte zu bestimmen. Obwohl es unendlich viele Belegungen v für A gibt, da jede Belegung die unendliche Menge der Atome auf je einen Wahrheitswert abbildet, ist leicht einzusehen, dass es genügt die Belegungen zu betrachten, die Atome in A mit verschiedenen Wahrheitswerten belegen. Die Auflistung aller relevanten Belegungen v, zusammen mit dem Wahrheitswert $\bar{\mathsf{v}}(A)$ wird *Wahrheitstabelle von* A genannt. Wenn A aus n verschiedenen Atomen zusammengesetzt ist, hat die Wahrheitstabelle für A maximal 2^n Zeilen, die wir prüfen müssen. Offensichtlich ist dieses Verfahren sehr einfach und bei kleinen Formeln auch recht schnell durchzuführen. Genauso offensichtlich aber ist die inhärente Komplexität.

2.3. Äquivalenz von Formeln

Definition 2.4: Äquivalenz

Zwei Formeln A, B sind *(logisch) äquivalent*, wenn $A \models B$ und $B \models A$ gilt. Wenn A und B äquivalent sind, dann schreiben wir kurz $A \equiv B$.

Satz 2.2

$A \equiv B$ gilt gdw. $(A \to B) \wedge (B \to A)$ eine Tautologie ist.

Beweis. Zunächst überlegt man sich leicht anhand der Wahrheitstabelle für $\wedge$, dass $(A \to B) \wedge (B \to A)$ eine Tautologie ist gdw. $(A \to B)$ und $(B \to A)$ Tautologien sind.

Nun betrachten wir die Annahme $A \models B$. Dann gilt für alle Belegungen v, dass $\bar{\mathsf{v}}(A) = \mathsf{T}$, $\bar{\mathsf{v}}(B) = \mathsf{T}$ impliziert. Somit gilt aber auch (kontrollieren Sie mit Hilfe der Wahrheitstabelle für $\to$), dass $\bar{\mathsf{v}}(A \to B) = \mathsf{T}$ für alle v. Also ist $A \to B$ eine Tautologie. Ähnlich folgt aus $B \models A$, dass $B \to A$ eine Tautologie ist. Somit haben wir gezeigt, dass $A \equiv B$ impliziert, dass $A \to B$ und $B \to A$ Tautologien sind. Mit Hilfe der anfänglichen Bemerkungen lässt sich der Satz von links nach rechts zeigen. Die Umkehrung folgt in der gleichen Weise und wird der Leserin überlassen. □

Die Konjunktion ist assoziativ, das heißt wir unterscheiden nicht zwischen $(A \wedge B) \wedge C$, $A \wedge (B \wedge C)$ und $A \wedge B \wedge C$. Statt $A_1 \wedge \cdots \wedge A_n$ schreiben wir auch $\bigwedge_{i=1}^{n} A_i$. Wenn $n = 0$, dann setzen wir $\bigwedge_{i=1}^{0} A_i := \mathsf{True}$. Das gleiche gilt für die Disjunktion und wir verwenden $\bigvee_{i=1}^{n} A_i$ für $A_1 \vee \cdots \vee A_n$ mit $\bigvee_{i=1}^{0} A_i := \mathsf{False}$. Zudem sind die Junktoren $\wedge$ und $\vee$ kommutativ. Das heißt $A \wedge B$ und $B \wedge A$ haben die gleiche Bedeutung.

Lemma 2.1

Es gelten die folgenden elementaren Äquivalenzen:

$$\begin{array}{llll} \neg\neg A \equiv A & A \vee \mathsf{True} \equiv \mathsf{True} & A \wedge \mathsf{True} \equiv A & A \to \mathsf{True} \equiv \mathsf{True} \\ & A \vee \mathsf{False} \equiv A & A \wedge \mathsf{False} \equiv \mathsf{False} & A \to \mathsf{False} \equiv \neg A \\ & A \vee A \equiv A & A \wedge A \equiv A & \mathsf{True} \to A \equiv A \\ & A \vee \neg A \equiv \mathsf{True} & A \wedge \neg A \equiv \mathsf{False} & \mathsf{False} \to A \equiv \mathsf{True} \\ & & & A \to A \equiv \mathsf{True} \end{array}$$

Lemma 2.2

Es gelten die folgenden Äquivalenzen zur Umformung verschiedener Junktoren:

$$\begin{array}{ll} A \to B \equiv \neg A \vee B & \neg(A \to B) \equiv A \wedge \neg B \\ A \wedge (B \vee C) \equiv (A \wedge B) \vee (A \wedge C) & A \vee (B \wedge C) \equiv (A \vee B) \wedge (A \vee C)\,. \end{array}$$

Die letzten beiden Äquivalenzen bedeuten, dass die *Distributivgesetze* für Konjunktion und Disjunktion gelten.

Lemma 2.3

Es gelten die folgenden *Absorptionsgesetze* zwischen Disjunktion und Konjunktion:

$$\begin{array}{ll} A \wedge (A \vee B) \equiv A & A \vee (A \wedge B) \equiv A \\ A \wedge (\neg A \vee B) \equiv A \wedge B & A \vee (\neg A \wedge B) \equiv A \vee B\,. \end{array}$$

Lemma 2.4

Es gelten die *Gesetze von de Morgan*:

$$\neg(A \wedge B) \equiv \neg A \vee \neg B \quad \neg(A \vee B) \equiv \neg A \wedge \neg B \,.$$

Alle oben angegebenen Äquivalenzen können durch das Aufstellen von Wahrheitstabellen nachgewiesen werden. Wir werden die Beweise für die Lemmata 2.1–2.4 im Rahmen von Booleschen Algebren in Kapitel 3 nachholen.

Eine *Teilformel* A einer Formel B ist ein Teilausdruck von B, der wiederum eine Formel ist. Den nächsten Satz geben wir ohne Beweis an, die interessierte Leserin wird an [5] verwiesen.

Satz 2.3

Sei A eine Formel und E eine Teilformel von A. Außerdem sei B eine Formel und F Teilformel von B. Gelte nun $E \equiv F$ und sei B das Resultat der Ersetzung von E durch F in A. Dann gilt $A \equiv B$. □

Sei A eine Formel und sei p ein Atom in A. Dann bezeichnet $A\{p \mapsto \mathsf{True}\}$ jene Formel, in welcher alle Vorkommnisse von p durch True ersetzt werden. Analog definiert man $A\{p \mapsto \mathsf{False}\}$.

Lemma 2.5

Sei A eine Formel und p ein Atom in A. Es gelten die folgenden Äquivalenzen:

1. A ist eine Tautologie gdw. $A\{p \mapsto \mathsf{True}\}$ und $A\{p \mapsto \mathsf{False}\}$ Tautologien sind.
2. A ist unerfüllbar gdw. $A\{p \mapsto \mathsf{True}\}$ und $A\{p \mapsto \mathsf{False}\}$ unerfüllbar sind.

Das obige Lemma liefert die Grundlage für eine Methode die Gültigkeit von Formeln zu überprüfen, welche *Methode von Quine* heißt [7]. Die Atome in der gegebenen Formel werden sukzessive durch True beziehungsweise False ersetzt, sodass grundlegende Äquivalenzen, wie die obigen, verwendbar werden. Diese Methode ist, im Gegensatz zur Wahrheitstabellenmethode, auch für größere Formeln verwendbar. Trotzdem bleibt sie (notwendigerweise) ineffizient. Die inhärente Komplexität des *Erfüllbarkeitsproblems*, werden wir noch näher im Kapitel 6 untersuchen.

2.4. Natürliches Schließen

Wir wenden uns nun syntaktischen Methoden zur Bestimmung der Gültigkeit einer Formel zu. Dazu definieren wir einen "passenden" *Kalkül* der Aussagenlogik. Im Allgemeinen spricht man von einem *Kalkül*, wenn eine fixe (formale) Sprache und Regeln zum Formen bestimmter Ausdrücke in dieser Sprache gegeben sind. Im Weiteren muss den Ausdrücken eine Bedeutung zuordenbar sein und es muss eindeutige Regeln geben, wie ein Ausdruck in einen anderen Ausdruck umgewandelt werden kann.

Für die Aussagenlogik betrachten wir dazu den Kalkül $\mathcal{NK}$ des *natürlichen Schließens*, der auf Georg Gentzen (1909–1945) zurückgeht [10]. Dieser Kalkül ist "passend", da alle verwendeten Regeln *korrekt* sind und alle Regeln gemeinsam *vollständig* sind. Das heißt, sämtliche in $\mathcal{NK}$ ableitbaren Formeln sind tatsächlich Tautologien („Korrektheit") und andererseits sind sämtliche Tautologien auch wirklich ableitbar („Vollständigkeit").

	Einführung	*Elimination*
$\wedge$	$\dfrac{A \quad B}{A \wedge B}\ \wedge\colon \mathsf{i}$	$\dfrac{A \wedge B}{A}\ \wedge\colon \mathsf{e} \qquad \dfrac{A \wedge B}{B}\ \wedge\colon \mathsf{e}$
$\vee$	$\dfrac{A}{A \vee B}\ \vee\colon \mathsf{i} \qquad \dfrac{B}{A \vee B}\ \vee\colon \mathsf{i}$	$\dfrac{A \vee B \qquad \boxed{\begin{array}{c}A\\ \vdots\\ C\end{array}} \qquad \boxed{\begin{array}{c}B\\ \vdots\\ C\end{array}}}{C}\ \vee\colon \mathsf{e}$
$\rightarrow$	$\dfrac{\boxed{\begin{array}{c}A\\ \vdots\\ B\end{array}}}{A \rightarrow B}\ \rightarrow\colon \mathsf{i}$	$\dfrac{A \quad A \rightarrow B}{B}\ \rightarrow\colon \mathsf{e}$
$\neg$	$\dfrac{\boxed{\begin{array}{c}A\\ \vdots\\ \mathsf{False}\end{array}}}{\neg A}\ \neg\colon \mathsf{i}$	$\dfrac{A \quad \neg A}{\mathsf{False}}\ \neg\colon \mathsf{e}$
False		$\dfrac{\mathsf{False}}{A}\ \mathsf{False}\colon \mathsf{e}$
$\neg\neg$		$\dfrac{\neg\neg A}{A}\ \neg\neg\colon \mathsf{e}$

Abbildung 2.2.: Natürliches Schließen

Definition 2.5: Kalkül $\mathcal{NK}$

Der Kalkül $\mathcal{NK}$ des *natürlichen Schließens* besteht aus den Beweisregeln in Abbildung 2.2.

Beachten Sie, dass jedem Junktor, sofern möglich, *Einführungs-* bzw. *Eliminationsregeln* zugeordnet werden. Zum Beispiel gibt es für die Konjunktion die Einführungsregel (mit $\wedge\colon \mathsf{i}$ bezeichnet) und zwei Eliminationsregeln (mit $\wedge\colon \mathsf{e}$).[1]

Definition 2.6: Ableitung

Sei $\mathcal{G}$ eine endliche Menge von Formeln und F eine Formel.

1. Ein *Beweis* von F aus $\mathcal{G}$ ist eine Folge von Formeln $A_1, \ldots, A_n$ mit $A_n = F$, sodass für $i = 1, \ldots, n$ gilt: entweder $A_i \in \mathcal{G}$ oder A_i folgt durch Anwendung einer der Regeln in Abbildung 2.2 aus bereits abgeleiteten Formeln $\{A_j\colon j < i\}$.

2. Eine Formel F heißt *beweisbar* aus den Annahmen $\mathcal{G}$, wenn es einen Beweis von F aus $\mathcal{G}$ gibt. Ein Beweis wird oft auch als *Ableitung*, *Herleitung* oder *Deduktion* bezeichnet.

Bei einigen Regeln in Tabelle 2.2 ($\vee\colon \mathsf{e}$, $\rightarrow\colon \mathsf{i}$ und $\neg\colon \mathsf{i}$ stehen in den Prämissen der Regeln um-

[1] Hier steht i für "introduction" und e für "eliminiation".

rahmte Ableitungen. Diese Rahmen werden *Boxen* genannt und bezeichnen den Wirkungsbereich von *lokalen Annahmen*. Betrachten wir etwa die Einführungsregel der Implikation ($\to$: i):

$$\frac{\boxed{\begin{array}{l} A \\ \vdots \\ B \end{array}}}{A \to B} \to\text{: i}$$

Beachten Sie, dass hier zunächst die Prämisse A angenommen (und die Box geöffnet wird) und dann—üblicherweise unter Verwendung der Prämisse A—die Formel B abgeleitet wird. Nun schließen wir die Box und leiten die Implikation $A \to B$ ab. Damit wird auch die Prämisse A *geschlossen*.

Eine geschlossene Prämisse gilt nur lokal in ihrer Box und somit auch nicht außerhalb. Sie darf also auch nicht außerhalb der Box als Annahme verwendet werden. Umgekehrt bedeutet das, dass die abgeleitete Formel (hier die Implikation $A \to B$) nicht von der Gültigkeit von A abhängt.

Definition 2.7: Beweisbarkeitsrelation

Die *Beweisbarkeitsrelation*

$$\{A_1, \ldots, A_n\} \vdash B \ .$$

zeigt an, dass B aus $A_1, \ldots, A_n$ beweisbar ist. Wir schreiben $\vdash A$ statt $\varnothing \vdash A$ und nennen A in diesem Fall *beweisbar*. Der Einfachheit halber schreiben wir oft $A_1, \ldots, A_n \vdash B$, statt $\{A_1, \ldots, A_n\} \vdash B$.

Den Ausdruck $A_1, \ldots, A_n \vdash B$ nennt man auch ein *Sequent*; das Sequent ist *gültig*, wenn $A_1, \ldots, A_n \vdash B$ gilt.

Ein System des formalen Beweisens ist *korrekt*, wenn aus $A_1, \ldots, A_n \vdash B$ auch $A_1, \ldots, A_n \models B$ folgt. Das Beweissystem heißt *vollständig*, wenn aus $A_1, \ldots, A_n \models B$ auch $A_1, \ldots, A_n \vdash B$ folgt. In einem korrekten und vollständigen Beweissystem ist die (syntaktische) Beweisbarkeitsrelation $\vdash$ der (semantischen) Konsequenzrelation $\models$ (Definition 2.3) äquivalent. Der Beweis des folgenden Satzes kann zum Beispiel in [10] nachgelesen werden.

Satz 2.4

Der Kalkül $\mathcal{NK}$ des natürlichen Schließens ist korrekt und vollständig für die Aussagenlogik: $A_1, \ldots, A_n \models B$ gdw. $A_1, \ldots, A_n \vdash B$. □

Der nächste Satz, das *Deduktionstheorem*, kann das formale Argumentieren erheblich erleichtern.

Satz 2.5

Wenn $A_1, \ldots, A_i, \ldots, A_n \vdash B$ gilt, dann auch $A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n \vdash A_i \to B$, das heißt, wenn A_i eine Prämisse in einem Beweis von B ist, dann existiert ein Beweis von $A_i \to B$, der A_i nicht als Prämisse hat.

Beweis. Nach Annahme gilt $A_1, \ldots, A_i, \ldots, A_n \vdash B$. OBdA.[2] können wir annehmen, dass $n = i$. Also gibt es einen Beweis in $\mathcal{NK}$ der folgenden Gestalt:

[2] Wir schreiben OBdA. als Abkürzung für „Ohne Beschränkung der Allgemeinheit“.

1	A_1	Prämisse
$\vdots$	$\vdots$	
$n-1$	A_{n-1}	Prämisse
n	A_n	Prämisse
$\vdots$	$\vdots$	
k	B	

Nun fügen wir diesem Beweis noch eine Anwendung der $\to\colon \mathsf{i}$ Regel auf die Formeln A_n und B hinzu, wodurch aus der Prämisse A_n eine (lokale) Annahme wird. Wir erhalten also einen Beweis von $A_n \to B$:

1	A_1	Prämisse
$\vdots$	$\vdots$	
$n-1$	A_{n-1}	Prämisse
n	A_n	Annahme
$\vdots$	$\vdots$	
k	B	
$k+1$	$A_n \to B$	$\to\colon \mathsf{i}$

Somit ist die Prämisse A_n aus der Liste der Annahmen eliminiert und wir haben im Allgemeinen die Korrektheit des folgenden Sequents nachgewiesen:

$$A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n \vdash A_i \to B\ . \qquad \square$$

Beachten Sie, dass das Deduktionstheorem einer üblichen Beweismethode für einen Satz der Form „Wenn H, dann K“ entspricht: Um die Implikation zu zeigen, nehmen wir zunächst H als Prämisse an und zeigen dann K, siehe Kapitel 1.

Der nächste Satz drückt das Prinzip des *Widerspruchsbeweises* aus. Die Ableitung dieser Inferenzregel überlassen wir der Leserin, siehe Aufgabe 2.12. Eine allgemeine Erklärung zu Beweisprinzipien—wie eben dem Widerspruchsbeweis—finden Sie im Teil I des Skriptums im Kapitel 1.

Satz 2.6

Wenn $A, \neg B \vdash \mathsf{False}$ gilt, dann gilt auch $A \vdash B$, das heißt, um die Aussage B aus A zu folgern, kann *indirekt* vorgegangen werden: die Negation von B wird als weitere Prämisse angenommen und daraus wird ein Widerspruch (False) abgeleitet. $\square$

Die oben angegebene Axiomatisierung ist nicht die einzige korrekte und vollständige Menge von Schlußregeln, die für die Aussagenlogik angegeben werden kann. Es existiert eine Vielzahl von Axiomen- und anderen Beweissystemen, die korrekt und vollständig sind. Als Beispiel wollen wir hier das auf *Gottlob Frege* (1848–1925) und *Jan Łukasiewicz* (1878–1956) zurückgehende Axiomensystem nennen, siehe [7].

Es soll auch nicht unerwähnt bleiben, dass Schlussfolgerungen in der Aussagenlogik nicht nur formalisiert werden können, sondern dass es auch Kalküle gibt, die es erlauben einen Beweis automatisch zu finden. Obwohl diese Kalküle im schlimmsten Fall genauso ineffizient sind wie die Wahrheitstabellenmethode, sind diese Methoden in der Praxis sehr schnell in der Lage die Gültigkeit einer Formel zu verifizieren [12]. Der Kalkül des natürlichen Schließens, sowie sogenante „SAT-solver“ werden in der Vorlesung „Logic in Computer Science“ vertieft.

Definition 2.8

Eine Logik heißt *endlich axiomatisierbar*, wenn es eine *endliche* Menge von Axiomen und Inferenzregeln gibt, die korrekt und vollständig für diese Logik ist.

2.5. Konjunktive und Disjunktive Normalform

Definition 2.9

Eine *Wahrheitsfunktion* $f\colon \{\mathsf{T},\mathsf{F}\}^n \to \{\mathsf{T},\mathsf{F}\}$ ist eine Funktion, die n Wahrheitswerten einen Wahrheitswert zuordnet.

Das folgende Lemma folgt unmittelbar aus der Definition.

Lemma 2.6

Zu jeder Formel A existiert eine Wahrheitsfunktion f, die mit Hilfe der Wahrheitstabelle von A definiert wird.

In diesem Abschnitt werden wir zeigen, dass auch jeder Wahrheitsfunktion in eindeutiger Weise eine Formel zugeordnet werden kann.

Definition 2.10

Sei $f\colon \{\mathsf{T},\mathsf{F}\}^n \to \{\mathsf{T},\mathsf{F}\}$ eine Wahrheitsfunktion. Wir definieren die Menge aller Argumentsequenzen $\mathsf{TV}(f)$, sodass die Wahrheitsfunktion f T liefert:

$$\mathsf{TV}(f) := \{(s_1,\ldots,s_n) \mid f(s_1,\ldots,s_n) = \mathsf{T}\}\ .$$

Definition 2.11: Konjunktive und Disjunktive Normalform

Sei A eine Formel.

1. Ein *Literal* ist ein Atom oder die Negation eines Atoms.
2. A ist in *disjunktiver Normalform* (kurz *DNF*), wenn A eine Disjunktion von Konjunktionen von Literalen ist.
3. A ist in *konjunktiver Normalform* (kurz *KNF*), wenn A eine Konjunktion von Disjunktionen von Literalen ist.

Lemma 2.7

Sei $f\colon \{\mathsf{T},\mathsf{F}\}^n \to \{\mathsf{T},\mathsf{F}\}$ eine Wahrheitsfunktion mit $\mathsf{TV}(f) \neq \varnothing$ und $\mathsf{TV}(f) \neq \{\mathsf{T},\mathsf{F}\}^n$. Im Weiteren seien $p_1,\ldots,p_n$ paarweise verschiedene Atome.

1. Wir definieren:

$$D := \bigvee_{(s_1,\ldots,s_n)\in\mathsf{TV}(f)} \bigwedge_{i=1}^{n} A_i\ ,$$

wobei $A_i = p_i$, wenn $s_i = \mathsf{T}$ und $A_i = \neg p_i$ sonst. Dann ist D eine DNF, deren Wahrheitstabelle der Funktion f entspricht.

2. Wir definieren:

$$K := \bigwedge_{(s_1,\ldots,s_n)\notin \mathsf{TV}(f)} \bigvee_{j=1}^{n} B_j \ ,$$

wobei $B_j = p_j$, wenn $s_j = \mathsf{F}$ und $B_j = \neg p_j$ sonst. Dann ist K eine KNF, deren Wahrheitstabelle der Funktion f entspricht.

Beweis. Zunächst betrachten wir die erste Behauptung: Jede Argumentfolge $\overline{s} = (s_1, \ldots, s_n)$ über $\{\mathsf{T}, \mathsf{F}\}$ induziert eine Belegung $\mathsf{v}_{\overline{s}}$ in eindeutiger Weise. Sei $A_{\overline{s}} = \bigwedge_{i=1}^{n} A_i$ eine Konjunktion in D, sodass $A_i = p_i$, wenn $s_i = \mathsf{T}$ und $A_i = \neg p_i$ sonst. Dann gilt $\bar{\mathsf{v}}(A_{\overline{s}}) = \mathsf{T}$ gdw. $\mathsf{v} = \mathsf{v}_{\overline{s}}$. Da D eine Disjunktion ist, die genau aus den Konjunktionen $A_{(s_1,\ldots,s_n)}$ mit $f(s_1, \ldots, s_n) = \mathsf{T}$ zusammengesetzt ist, gilt:

$$\bar{\mathsf{v}}(D) = \mathsf{T} \qquad \text{gdw.} \qquad \mathsf{v} = \mathsf{v}_{(s_1,\ldots,s_n)} \quad \text{und} \quad f(s_1, \ldots, s_n) = \mathsf{T} \ .$$

Somit stimmen die Wahrheitstabelle von D und die Funktion f überein.

Im Falle der zweiten Behauptung betrachte man eine Disjunktion $B_{\overline{s}}$ in K. Es gilt: $\bar{\mathsf{v}}(B_{\overline{s}}) = \mathsf{F}$ gdw. $\mathsf{v} = \mathsf{v}_{\overline{s}}$. Somit gilt:

$$\bar{\mathsf{v}}(K) = \mathsf{F} \qquad \text{gdw.} \qquad \mathsf{v} = \mathsf{v}_{(s_1,\ldots,s_n)} \quad \text{und} \quad f(s_1, \ldots, s_n) = \mathsf{F} \ .$$

Also stimmen die Wahrheitstabelle von K und die Funktion f überein. □

Der nächste Satz folgt aus den Lemmata 2.6 und 2.7.

Satz 2.7

Jeder nicht-trivialen Wahrheitsfunktion ist auf eine eindeutige Weise eine DNF und eine KNF zugeordnet. Umgekehrt induziert jede Formel mit n Atomen eine eindeutige Wahrheitsfunktion in n Variablen.

Folgerung 2.1

Für jede Formel A existiert eine DNF D und eine KNF K, sodass $A \equiv D \equiv K$ gilt.

Beweis. Es genügt auf die Fälle einzugehen, in denen A eine Tautologie oder unerfüllbar ist. Wenn A eine Tautologie, dann setzen wir $D = K := p \vee \neg p$. Andererseits, wenn A unerfüllbar ist, dann setzen wir $D = K := p \wedge \neg p$, wobei p ein beliebiges Atom ist. □

Alternativ lassen sich DNF und KNF durch elementare Umformungen erhalten [10].

2.6. Zusammenfassung

Obwohl die Logik ursprünglich eine philosophische Disziplin mit einer starken mathematisch-formalen Komponente ist, hat sich die (mathematische) Logik in den letzten Dekaden als die

entscheidende Grundlagendisziplin der Informatik herausgestellt. Dies ist leicht erklärbar: Die Hauptaufgabe einer Informatikerin ist es, eine informelle Beschreibung eines Problems (eine *Spezifikation*) durch Abstraktion so umwandeln zu können, dass dieses Problem in einer *formalen Sprache* (also mit einem Programm) gelöst werden kann. In der Logik wurden (über Jahrhunderte) Methoden und Werkzeuge untersucht, die es erlauben diesen Übergang leicht und korrekt durchzuführen.

Die hier betrachtete Aussagenlogik behandelt nur Aussagen, die entweder wahr oder falsch sind. Das heißt die Logik ist *zweiwertig*, da wir genau zwei Wahrheitswerte (T und F) haben und es gilt das *Gesetz des ausgeschlossenen Dritten*: Entweder gilt eine Aussage A oder ihr Gegenteil $\neg A$. Anders ausgedrückt: $A \vee \neg A$ ist eine Tautologie. Wir können auch Logiken betrachten, für die dies nicht gilt. Einerseits gibt es Logiken, die das Gesetz des ausgeschlossenen Dritten nicht beinhalten. Solche Logiken nennt man *intuitionistisch*. Der zentrale Unterschied zwischen einer intuitionistischen Aussagenlogik und der Aussagenlogik, die wir in diesem Kapitel behandelt haben, ist, dass erstere *indirekte* Beweise (wie in Satz 2.6) nicht erlaubt: jeder Beweis muss direkt sein. Andererseits gibt es Logiken mit mehr als zwei Wahrheitswerten. Diese Logiken nennt man *mehrwertig*.

Wir gehen kurz auf ein Beispiel und eine Anwendung von mehrwertigen Logiken ein. Sei $V \subseteq [0,1]$ eine Menge von Wahrheitswerten, sodass V zumindest die Werte 0 und 1 enthält. Hier steht 0 für eine zweifelsfrei falsche Aussage und 1 für eine zweifelsfrei richtige Aussage. Eine *Łukasiewicz-Belegung* (basierend auf V) ist eine Abbildung $\mathsf{v}\colon \mathsf{AT} \to V$ und diese Belegung wird wie folgt zu einem *Wahrheitswert* erweitert:

$$\begin{aligned}
\bar{\mathsf{v}}(\neg A) &= 1 - \bar{\mathsf{v}}(A) \\
\bar{\mathsf{v}}(A \wedge B) &= \min\{\bar{\mathsf{v}}(A), \bar{\mathsf{v}}(B)\} \\
\bar{\mathsf{v}}(A \vee B) &= \max\{\bar{\mathsf{v}}(A), \bar{\mathsf{v}}(B)\} \\
\bar{\mathsf{v}}(A \to B) &= \max\{1, 1 - \bar{\mathsf{v}}(A) + \bar{\mathsf{v}}(B)\}
\end{aligned}$$

Eine Formel A heißt *gültig*, wenn $\bar{\mathsf{v}}(A) = 1$ für alle Łukasiewicz-Belegungen v.

Mehrwertige Logiken, die auf einer Łukasiewicz-Belegung aufbauen, werden Łukasiewicz-Logiken genannt. Manchmal werden solche Logiken auch *Fuzzy*-Logiken genannt. Obwohl diese Logiken unendlich viele Wahrheitswerte verwenden können, sind sie endlich axiomatisierbar. Die Bedeutung solcher mehrwertiger Logiken in der Informatik wird durch die Möglichkeit gegeben, Unsicherheit von Information auszudrücken. Etwa finden Łukasiewicz-Logiken praktische Anwendung in medizinischen Expertensystemen.

2.7. Aufgaben

Übungsaufgabe 2.1

Wie sind die Begriffe *Vereingung*, *Durchschnitt* und *Differenz* von *Mengen* definiert?
Prüfen Sie nach, ob für beliebige Mengen A, B und C die folgenden Aussagen allgemeingültig sind:

1. $A \setminus (B \cup C) = (A \setminus B) \cup (A \setminus C)$
2. $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$

Hinweis: Studieren Sie die „Formalen Konzepte“ des Brückenkurses.

Übungsaufgabe 2.2

Zwei Rotmützen- und zwei Blaumützenzwerge sind unterwegs im Logikland. Unglücklicherweise werden die Zwerge von einem Gnom gefangen, der droht, sie seinem Drachen zum Fraß vorzuwerfen. Die Zwerge bekommen jedoch eine letzte Chance.
Der Gnom stellt einen der Zwerge vor und drei hinter eine hohe Mauer und vertauscht (in der Nacht) folgendermaßen deren Mützen:

$$\mathrm{R} \mid \mathrm{B} \quad \mathrm{R} \quad \mathrm{B}$$

(R bezeichnet eine rote, B eine blaue Mütze, alle blicken zur Mauer.) Falls einer der Zwerge innerhalb eines Tages die Farbe seiner Mütze errät, kommen sie frei. (Sie dürfen sich natürlich nicht bewegen, und nicht miteinander sprechen.)
Welcher Zwerg errät seine Mützenfarbe?

Lösung. Der zweite Zwerg rechts von der Mauer kann seine Mützenfarbe wie folgt erkennen: Entweder habe ich eine rote oder eine blaue Mütze. Angenommen ich habe eine blaue Mütze, dann habe ich und der Zwerg vor mir eine blaue Mütze, also muss der Zwerg hinter mir eine rote Mütze haben und dies auch ausdrücken. Nun bleibt der Zwerg jedoch stumm, also muss ich eine rote Mütze haben.

Übungsaufgabe 2.3

Welche der folgenden Schlussfiguren sind korrekt?

1. *Sokrates ist ein Mensch.*
   Alle Menschen sind Philosophen.
   Sokrates ist ein Philosoph.

2. *Sokrates ist ein Mensch.*
   Alle Griechen sind reich.
   Sokrates ist reich.

3. *Sokrates ist ein Mensch.*
   Alle Menschen sind sterblich.
   Sokrates ist ein Lebewesen.

A gilt.
4. *B gilt nicht.*
Wenn B gilt, dann gilt A.

Lösung.

1. Korrekt, da aus den Prämissen die Konklusion folgt (auch wenn die zweite Prämisse falsch ist).
2. Korrekt, da die zweite Prämisse falsch ist. Wenn alle Griechen reich wären, dann wäre der Syllogismus falsch, weil wir zu wenig über Sokrates wissen (er muss kein Grieche sein).
3. Korrekt (da natürlich jeder Mensch ein Lebewesen ist). Diese Information haben wir in unserem Schluss aber nicht zur Verfügung, somit ist dies kein üblicher Syllogismus. Um dieses Phänomen zukünftig auszuschließen werden wir uns ab Woche 2 nur mehr mit abstrakten Aussagen A, B, etc. beschäftigen.
4. Korrekt, da A eine Prämisse ist und laut Annahme immer gilt.

Übungsaufgabe 2.4

Formalisieren Sie folgende Sätze über den Straßenverkehr als propositionale Formeln (Beispiel: Ein blaues Auto folgt auf ein schwarzes Auto. $S \to B$):

1. Kommen ein rotes Auto und ein gelbes Auto, so folgt ein oranges Auto.
2. Es kommt ein oranges Auto oder nach jedem roten Auto folgt ein gelbes Auto.
3. Es kommt ein rotes oder gelbes Auto oder kein oranges Auto.
4. Nach jedem roten Auto kommt ein gelbes Auto oder es kommt kein rotes Auto.

Lösung.

1. $(A \wedge B) \to C$
2. $C \vee (A \to B)$
3. $(A \vee B \vee \neg C)$
4. $A \to B \vee \neg A$

Übungsaufgabe 2.5

Betrachten Sie die elementaren Äquivalenzen in den Lemmmata 2.1–2.4 und zeigen Sie jeweils 1–2 Äquivalenzen für jedes Lemma mit der Methode der Wahrheitstabelle.

Übungsaufgabe 2.6

Welche der folgenden Formeln sind (i) erfüllbar, (ii) unerfüllbar, (iii) gültig?

1. $p \to p$
2. $q \to \neg(r \wedge p)$
3. $q \vee (q \to (p \wedge \neg p))$

Hinweis: Argumentieren Sie mittels Wahrheitstabellen. Können Sie für die gültigen Formeln A mithilfe der Gesetze aus der Vorlesung $A \equiv \mathsf{True}$ beweisen?

Übungsaufgabe 2.7

Welche der folgenden Konsequenzrelationen gelten? Welche sind Äquivalenzen?

1. $p \wedge (q \wedge r) \models (p \wedge r) \wedge q$
2. $p \to (q \to r) \models (p \to q) \to r$
3. $p \wedge \neg p \models q$

Hinweis: Argumentieren Sie mittels Wahrheitstabellen. Können Sie die Äquivalenzen mithilfe der Gesetze aus der Vorlesung beweisen?

Übungsaufgabe 2.8

Welche der folgenden Aussagen sind wahr?

1. Wenn eine Formel A gültig ist, dann ist A erfüllbar.
2. Wenn eine Formel A erfüllbar ist, dann ist A gültig.
3. Wenn eine Formel A erfüllbar ist, dann ist $\neg A$ unerfüllbar.

Hinweis: Geben Sie Beweise für die wahren Aussagen und Gegenbeispiele für die falschen.

Übungsaufgabe 2.9

Prüfen Sie folgende Formeln mit Hilfe der *Methode von Quine* auf die Eigenschaften Unerfüllbarkeit sowie Tautologie.

1. $p \to (q \to p)$
2. $(p \to (q \to r)) \to ((p \to q) \to (p \to r))$
3. $(p \wedge q) \wedge \neg(p \wedge q)$

Lösung.

1. Wie man in folgendem Baum sehen kann, sind sowohl

$$(p \to (q \to p))\{p \mapsto \mathsf{True}\}$$

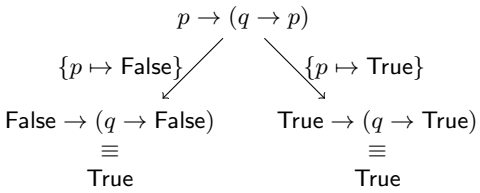
als auch

$$(p \to (q \to p))\{p \mapsto \mathsf{False}\}$$

Tautologien und somit ist nach Lemma 1.5 im Skript auch

$$p \to (q \to p)$$

eine Tautologie.



2. Wie man in folgendem Baum sehen kann, sind sowohl

$$((p \to (q \to r)) \to ((p \to q) \to (p \to r)))\{p \mapsto \mathsf{True}\}$$

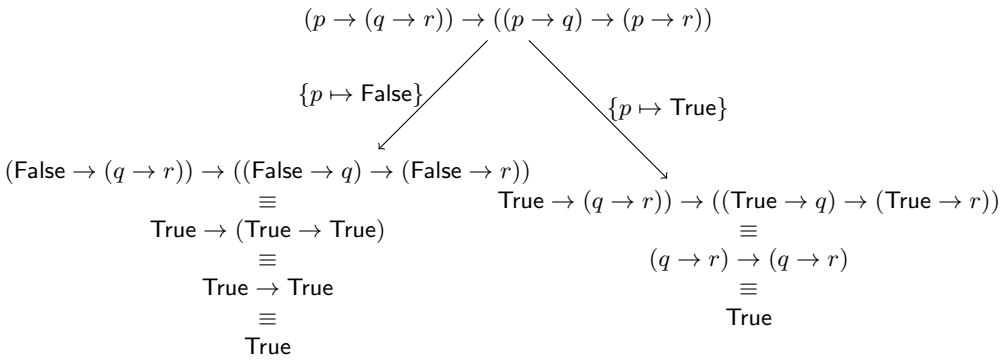
als auch

$$((p \to (q \to r)) \to ((p \to q) \to (p \to r)))\{p \mapsto \mathsf{False}\}$$

Tautologien und somit ist nach Lemma 1.5 im Skriptum auch

$$(p \to (q \to r)) \to ((p \to q) \to (p \to r))$$

eine Tautologie.



Übungsaufgabe 2.10

Beweisen Sie $\vdash p \to p$ im Kalkül $\mathcal{NK}$.

Lösung. Die erste und letzte Zeile einer Box können auch dieselbe sein.

1	p	
2	$p \to p$	$\to$: i 1–1

Übungsaufgabe 2.11

Die folgende Operation ($\overline{\wedge}$) wird NAND (engl. für „negated and“) genannt.

p	q	$p \mathbin{\overline{\wedge}} q$
F	F	T
F	T	T
T	F	T
T	T	F

Erstellen Sie die

1. konjunktive Normalform
2. disjunktive Normalform

von NAND, indem Sie sich genau an die in Lemma 1.7 vorgestellte Methode halten, das heißt, stellen Sie $\mathsf{TV}(\overline{\wedge})$ auf und leiten Sie davon die Normalformen ab.

Übungsaufgabe 2.12

Weisen Sie die folgende Inferenzregeln als ableitbar im Kalkül $\mathcal{NK}$ nach:

$$\dfrac{\boxed{\begin{array}{l}\neg A\\ \vdots\\ \mathsf{False}\end{array}}}{A}$$

Lösung. Um dies zu zeigen, nehmen wir an, dass man in einer Box bereits False von $\neg A$ abgeleitet hat. Wir müssen dann nur zeigen, dass wir davon A ableiten können.

1	$\neg A$	
$\vdots$	$\vdots$	
n	False	
$n+1$	$\neg\neg A$	$\neg$: i 1–n
$n+2$	A	$\neg\neg$: e $n+1$

Übungsaufgabe 2.13

Zeigen Sie mit Hilfe der Methode von Quine, dass die folgenden Formeln Tautologien sind:

1. $\neg p \to (p \to q)$
2. $p \to (\neg p \to q)$

Lösung.

Übungsaufgabe 2.14

Zeigen Sie die Gültigkeit der Formeln (i) $\neg p \to (p \to q)$; (ii) $p \to (\neg p \to q)$ (siehe Aufgabe 2.13) in $\mathcal{NK}$.

Lösung.

(i)

1	$\neg p$	
2	p	
3	False	False: i 1,2
4	q	False: e 3
5	$p \to q$	$\to$: i 2–4
6	$\neg p \to (p \to q)$	$\to$: i 1–5

(ii)

1	p	
2	$\neg p$	
3	False	False: i 1,2
4	q	False: e 3
5	$\neg p \to q$	$\to$: i 2–4
6	$p \to (\neg p \to q)$	$\to$: i 1–5

Übungsaufgabe 2.15

Betrachten Sie die beiden Formeln in 2.13 und 2.14. Beweisen oder widerlegen Sie die folgende Behauptung: „Nur, wenn wir beide Aufgaben gelöst haben, dann können wir sicher sein, dass diese Formeln sowohl Tautologien als auch allgemein gültig sind.“

Lösung. Die Behauptung ist falsch. Wenn wir uns nur für die Gültigkeit der Formeln interessieren können wir entweder syntaktisch (im $\mathcal{NK}$) oder semantisch (mit Wahrheitstafeln oder Quine) argumentieren. Das folgt aus Satz 2.4.

Übungsaufgabe 2.16

Lesen Sie den XKCD comic 468 (`http://xkcd.com/468/`).
Hinweis: Eventuell können Sie erst am Ende Ihres Studiums lachen.

3.

Einführung in die Algebra

In diesem Kapitel werden *Algebren* und im Besonderen *Boolesche Algebren* eingeführt. Zum leichteren Verständnis der Begrifflichkeiten werden zunächst in Abschnitt 3.1 allgemeine Sachverhalte zu *Algebren* besprochen, wie sie auch in der LVA „Lineare Algebra“ behandelt werden. Im Kapitel 3.2 wenden wir uns dann Booleschen Algebren zu. Boolesche Algebren finden etwa in der Konstruktion von Schaltkreisen ihre Verwendung, deshalb wird diese Anwendung auch im Kapitel A.1 (im Anhang) vertieft, siehe auch [8].

Schließlich gehen wir in Abschnitt 3.3 kurz auf die Bedeutung der Algebra für die Informatik ein und stellen die betrachtete Konzepte in einen historischen Kontext. Außerdem finden sich in Abschnitt 3.4 (optionale) Aufgaben zu den Themenbereichen dieses Kapitels, die zur weiteren Vertiefung dienen sollen.

3.1. Algebraische Strukturen

Algebren erlauben eine abstrakte Beschreibung von Objekten, indem die Eigenschaften dieser Objekte durch die Operationen, die mit diesen Objekten möglich sind, beschrieben werden.

Definition 3.1: Algebra

Eine *Algebra* $\mathcal{A} = \langle A_1, \ldots, A_n; \circ_1, \ldots, \circ_m \rangle$ ist eine Struktur, die aus den Mengen $A_1, \ldots, A_n$ und den Operationen $\circ_1, \ldots, \circ_m$ auf diesen Mengen besteht. Die Mengen $A_1, \ldots, A_n$ werden *Träger* (oder auch *Trägermengen*) genannt und nullstellige Operationen nennt man *Konstanten*.

Definition 3.2: Algebraischer Ausdruck

Sei $\mathcal{A}$ eine Algebra über den Trägermengen $A_1, \ldots, A_n$ und sei eine unendliche Menge von Variablen $x_1, x_2, \ldots$ gegeben, die als Platzhalter für Objekte in $A_1, \ldots, A_n$ verwendet werden. Im weiteren setzen wir für jede Operation $\circ_i$ der Algebra $\mathcal{A}$ ein Symbol der gleichen Stelligkeit voraus. Der Einfachheit halber bezeichnen wir dieses Symbol wieder mit $\circ_i$. Wir definieren die *algebraischen Ausdrücke von* $\mathcal{A}$ induktiv:

1. Konstanten und Variablen sind algebraische Ausdrücke.
2. Wenn $\circ$ eine Operation von $\mathcal{A}$ ist, die m Argumente hat und $E_1, \ldots, E_m$ algebraische Ausdrücke sind, dann ist $\circ(E_1, \ldots, E_m)$ ein algebraischer Ausdruck.

Konvention

Wann immer möglich schreiben wir Operationen in Infixnotation, zum Beispiel schrieben wir $a_1 \circ a_2$ statt $\circ(a_1, a_2)$ bei einer zweistelligen Operation $\circ$.

Algebraische Ausdrücke spielen eine ähnliche Rolle wie Formeln der Aussagenlogik. Sie stellen eine textuelle Beschreibung bestimmter Objekte der Trägermengen dar.

Definition 3.3

Sei $\mathcal{A}$ eine Algebra und seien E und F algebraische Ausdrücke über der Algebra $\mathcal{A}$. E' ist eine *Instanz* von E, wenn wir alle Variablen durch Elemente aus dem Träger von $\mathcal{A}$ ersetzen. Die Ausdrücke E und F nennen wir *äquivalent (in $\mathcal{A}$)*, wenn alle Instanzen von E und F (wobei Variablen in E und F in gleicher Weise durch Elemente aus dem Träger von $\mathcal{A}$ ersetzt werden) nach Auswertung in der Algebra $\mathcal{A}$ den selben Wert annehmen. Wenn E äquivalent zu F ist, schreiben wir kurz $E \approx F$.

Sei $\mathcal{A}$ eine Algebra mit endlichen Trägern $A_1, \ldots, A_n$. Dann nennen wir $\mathcal{A}$ *endlich*. Für endliche Algebren können die Operationen anhand einer *Operationstabelle*, die die Ergebnisse der Operationen auf den Trägerelementen angibt, festgelegt werden.

Definition 3.4

Sei $\circ$ eine binäre Operation auf der Menge A.

- Wenn $0 \in A$ existiert, sodass für alle $a \in A$: $a \circ 0 = 0 \circ a = 0$, dann heißt 0 *Nullelement* für $\circ$.
- Wenn $1 \in A$ existiert, sodass für alle $a \in A$: $a \circ 1 = 1 \circ a = a$, dann heißt 1 *Einselement* oder *neutrales Element* für $\circ$.
- Sei 1 das neutrale Element für $\circ$ und angenommen für $a \in A$, existiert $b \in A$, sodass $a \circ b = b \circ a = 1$. Dann heißt b das *Inverse* oder das *Komplement* von a.

Lemma 3.1

Jede binäre Operation hat maximal ein neutrales Element.

Beweis. Sei $\circ$ eine binäre Operation auf der Menge A und angenommen e und u sind neutrale Elemente für $\circ$. Wir zeigen, dass $e = u$. Somit kann es nur ein neutrales Element geben.

$$\begin{aligned} e &= e \circ u && \text{da } u \text{ neutrales Element} \\ &= u && \text{da } e \text{ neutrales Element .} \end{aligned}$$

□

Definition 3.5

Sei $\mathcal{A} = \langle A; \circ \rangle$ eine Algebra. Dann heißt

- $\langle A; \circ \rangle$ *Halbgruppe*, wenn $\circ$ assoziativ ist.

- $\langle A; \circ, 1\rangle$ *Monoid*, wenn $\langle A; \circ\rangle$ eine Halbgruppe ist und 1 ein neutrales Element für $\circ$ ist.
- $\langle A; \circ, 1\rangle$ *Gruppe*, wenn $\langle A; \circ, 1\rangle$ ein Monoid ist und jedes Element ein Inverses besitzt.

Eine Halbgruppe, ein Monoid oder eine Gruppe heißt *kommutativ*, wenn $\circ$ kommutativ ist.

Lemma 3.2

Wenn $\mathcal{A} = \langle A; \circ, 1\rangle$ ein Monoid ist, dann ist das Inverse eindeutig.

Beweis. Sei $a \in A$ und seien b, c Inverse von a. Wir zeigen $b = c$.

$$\begin{aligned} b &= b \circ 1 && 1 \text{ ist neutrales Element} \\ &= b \circ (a \circ c) && c \text{ ist Inverses von } a \\ &= (b \circ a) \circ c && \text{Assoziativität von } \circ \\ &= 1 \circ c && b \text{ ist Inverses von } a \\ &= c && 1 \text{ ist neutrales Element .} \end{aligned}$$

□

Definition 3.6

Sei $\mathcal{A} = \langle A; +, \cdot, 0, 1\rangle$ eine Algebra.

- $\mathcal{A}$ heißt *Ring*, wenn $\langle A; +, 0\rangle$ eine kommutative Gruppe ist und $\langle A; \cdot, 1\rangle$ ein Monoid. Im Weiteren muss gelten, dass $\cdot$ über $+$ distribuiert und zwar sowohl von links als auch von rechts. Das heißt für alle $a, b, c \in A$ gilt:

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c) \qquad (b + c) \cdot a = (b \cdot a) + (c \cdot a) \; .$$

- Wenn $\mathcal{A}$ ein kommutativer Ring ist, der nullteilerfrei ist, dann heißt $\mathcal{A}$ *Integritätsbereich*. Nullteilerfrei bedeutet, dass es keine Elemente $a, b \in A$ gibt, sodass $a \cdot b = 0$, aber $a \neq 0$ und $b \neq 0$.
- Wenn $\mathcal{A}$ ein Ring ist und darüber hinaus $\langle A \setminus \{0\}; \cdot, 1\rangle$ eine kommutative Gruppe, dann heißt $\mathcal{A}$ ein *Körper*.

Wegen der Nullteilerfreiheit von Integritätsbereichen verallgemeinern diese die ganzen Zahlen $\mathbb{Z}$ mit den üblichen Operationen. Andererseits ist $\mathbb{Z}$ kein Körper. Es gilt aber, dass jeder endliche Integritätsbereich ein Körper ist.

3.2. Boolesche Algebra

Definition 3.7: Boolesche Algebra

Eine Algebra $\mathcal{B} = \langle B; +, \cdot, \sim, 0, 1\rangle$ heißt *Boolesche Algebra* wenn gilt:

1. $\langle B; +, 0\rangle$ und $\langle B; \cdot, 1\rangle$ sind kommutative Monoide.
2. Die Operationen $+$ und $\cdot$ distribuieren übereinander. Es gilt also für alle $a, b, c \in B$:
$$a \cdot (b + c) = (a \cdot b) + (a \cdot c) \qquad a + (b \cdot c) = (a + b) \cdot (a + c)\ .$$
3. Für alle $a \in B$ gilt $a + \sim(a) = 1$ und $a \cdot \sim(a) = 0$. Das Element $\sim(a)$ heißt das *Komplement* oder die *Negation* von a.

Konvention

Zur Verbesserung der Lesbarkeit lassen wir das Zeichen $\cdot$ oft weg und schreiben ab statt $a \cdot b$. Außerdem vereinbaren wir, dass der Operator $\cdot$ stärker bindet als $+$.

Für Algebren haben wir die Sprache der algebraischen Ausdrücke eingeführt. Spezialisiert auf Boolesche Algebren bezeichnet man solche Ausdrücke als *Boolesche Ausdrücke*.

Definition 3.8: Boolescher Ausdruck

Sei eine unendliche Menge von Variablen $x_1, x_2, \ldots$ gegeben, die als Platzhalter für Objekte einer Booleschen Algebra verwendet werden. Diese Variablen heißen *Boolesche Variablen*. Wir definieren *Boolesche Ausdrücke* induktiv:

1. $\mathsf{0}$, $\mathsf{1}$ und Variablen sind Boolesche Ausdrücke.
2. Wenn E und F Boolesche Ausdrücke sind, dann sind
$$\sim(E) \qquad (E \cdot F) \qquad (E + F)\ ,$$
Boolesche Ausdrücke.

In der Folge definieren wir eine Reihe von Booleschen Algebren, die in vielfacher Weise Anwendung finden.

Definition 3.9

Sei M eine Menge. Wir bezeichnen mit $\mathcal{P}(M)$ die *Potenzmenge* von M, also $\mathcal{P}(M) := \{N \mid N \subseteq M\}$. Wir betrachten die Algebra $\langle \mathcal{P}(M); \cup, \cap, \sim, \varnothing, M\rangle$ wobei $\cup$ die Mengenvereinigung, $\cap$ die Schnittmenge, $\sim$ die Komplementärmenge und $\varnothing$ die leere Menge bezeichnet. Diese Algebra nennt man *Mengenalgebra*.

Lemma 3.3

Die Mengenalgebra ist eine Boolesche Algebra.

Beweis. Es lässt sich leicht nachprüfen, dass alle Axiome der Booleschen Algebra, wie in Definition 3.7 definiert, erfüllt sind. □

Definition 3.10

Sei $\mathbb{B} := \{0, 1\}$, wobei $0, 1 \in \mathbb{N}$. Wir betrachten die Algebra $\langle \mathbb{B}; +, \cdot, \sim, 0, 1 \rangle$, wobei die Operationen $+, \cdot, \sim$ wie in Abbildung 3.1 definiert sind. Die Algebra $\langle \mathbb{B}; +, \cdot, \sim, 0, 1 \rangle$ nennt man *binäre Algebra*.

$+$	1	0
1	1	1
0	1	0

$\cdot$	1	0
1	1	0
0	0	0

$\sim$	
1	0
0	1

Abbildung 3.1.: Operationen auf $\mathbb{B}$

Lemma 3.4

Die binäre Algebra ist eine Boolesche Algebra.

Beweis. Es lässt sich leicht nachprüfen, dass alle Axiome der Booleschen Algebra, wie in Definition 3.7 definiert, erfüllt sind. □

Ein Vergleich von Abbildung 3.1 mit den in Abbildung 2.1 definierten Junktoren legt einen Zusammenhang von Booleschen Algebren und der Aussagenlogik nahe. Dabei entspricht das Komplement der Negation, die Operation $\cdot$ der Konjunktion und $+$ der Disjunktion. Darüber hinaus werden die Zahlen 0 und 1 als die Wahrheitswerte F und T interpretiert. In dieser Interpretation spricht man auch von der *Algebra der Wahrheitswerte*.

Umgekehrt können wir die Menge der aussagenlogischen Formeln zusammen mit den in Kapitel 2 definierten Junktoren als Algebra betrachten.

Definition 3.11

Sei Frm die Menge der aussagenlogischen Formeln. Wir betrachten die Algebra $\langle \mathsf{Frm}; \vee, \wedge, \neg, \mathsf{False}, \mathsf{True} \rangle$.
In der Algebra $\langle \mathsf{Frm}; \vee, \wedge, \neg, \mathsf{False}, \mathsf{True} \rangle$ entspricht die Gleichheit von Booleschen Ausdrücken der logischen Äquivalenz.

Lemma 3.5

Die in Definition 3.11 definierte Algebra ist eine Boolesche Algebra.

Beweis. Es lässt sich leicht nachprüfen, dass alle Axiome der Booleschen Algebra, wie in Definition 3.7 definiert, erfüllt sind. □

Definition 3.12

Sei n eine natürliche Zahl und sei $\mathbb{B}^n$ das n-fache kartesische Produkt von $\mathbb{B}$, also $\mathbb{B}^n := \{(a_1, \ldots, a_n) \mid a_i \in \mathbb{B}\}$. Wir betrachten die Algebra

$$\langle \mathbb{B}^n; +, \cdot, \sim, (0, \ldots, 0), (1, \ldots, 1) \rangle \ ,$$

wobei die Operationen $+$, $\cdot$, $\sim$ als die komponentenweise Erweiterung der Operationen in Definition 3.10 definiert sind:

$$\begin{aligned}(a_1, \ldots, a_n) + (b_1, \ldots, b_n) &:= (a_1 + b_1, \ldots, a_n + b_n) \\ (a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) &:= (a_1 \cdot b_1, \ldots, a_n \cdot b_n) \\ \sim((a_1, \ldots, a_n)) &:= (\sim(a_1), \ldots, \sim(a_n)) \ .\end{aligned}$$

Lemma 3.6

Die in Definition 3.12 definierte Algebra ist eine Boolesche Algebra.

Beweis. Es lässt sich leicht nachprüfen, dass alle Axiome der Booleschen Algebra, wie in Definition 3.7 definiert, erfüllt sind. □

Definition 3.13

Seien n, m natürliche Zahlen. Sei Abb die Menge der Abbildungen von $\mathbb{B}^n$ nach $\mathbb{B}^m$. Wir betrachten die Algebra

$$\langle \mathsf{Abb}; +, \cdot, \sim, (\mathbf{0}, \ldots, \mathbf{0}), (\mathbf{1}, \ldots, \mathbf{1}) \rangle \ ,$$

wobei $(\mathbf{0}, \ldots, \mathbf{0})$ und $(\mathbf{1}, \ldots, \mathbf{1})$ konstante Funktionen auf $\mathbb{B}^n \to \mathbb{B}^m$ bezeichnen:

$$\begin{aligned}(\mathbf{0}, \ldots, \mathbf{0}) &\colon \mathbb{B}^n \to \mathbb{B}^m, (a_1, \ldots, a_n) \mapsto (\underbrace{0, \ldots, 0}_{m}) \\ (\mathbf{1}, \ldots, \mathbf{1}) &\colon \mathbb{B}^n \to \mathbb{B}^m, (a_1, \ldots, a_n) \mapsto (\underbrace{1, \ldots, 1}_{m}) \ .\end{aligned}$$

Wir betrachten die in Definition 3.12 eingeführte Algebra mit Trägermenge $\mathbb{B}^m$. Die dort eingeführten Operationen werden punktweise erweitert:

$$\begin{aligned}(f + g)(a_1, \ldots, a_n) &:= f(a_1, \ldots, a_n) + g(a_1, \ldots, a_n) \\ (f \cdot g)(a_1, \ldots, a_n) &:= f(a_1, \ldots, a_n) \cdot g(a_1, \ldots, a_n) \\ \sim(f)(a_1, \ldots, a_n) &:= \sim(f(a_1, \ldots, a_n)) \ .\end{aligned}$$

Diese Algebra nennt man die *Algebra der n-stelligen Booleschen Funktionen.*

Lemma 3.7

Die Algebra der n-stelligen Booleschen Funktionen ist eine Boolesche Algebra.

Beweis. Es lässt sich leicht nachprüfen, dass alle Axiome der Booleschen Algebra, wie in Definition 3.7 definiert, erfüllt sind. □

Aufbauend auf den Axiomen einer Booleschen Algebra gelten eine Reihe von Gleichheiten, die den in Kapitel 2 studierten logischen Äquivalenzen entsprechen. Diese werden in der Folge betrachtet. Wenn die Beweise leicht nachvollziehbar sind, überlassen wir diese der Leserin. Für Boolesche Algebren gilt das *Dualitätsprinzip*. Sei $\mathcal{B}$ eine Boolesche Algebra und gelte die Gleichheit A für $\mathcal{B}$, dann gilt eine entsprechende Gleichheit A' bei der alle Vorkommnisse von $+$ durch $\cdot$ (und umgekehrt) ersetzt werden sowie 0 und 1 vertauscht werden.

Lemma 3.8

Sei $\mathcal{B}$ eine Boolesche Algebra und sei B die Trägermenge von $\mathcal{B}$. Für alle $a \in B$ gelten die folgenden *Idempotenzgesetze*:

$$a \cdot a = a \qquad a + a = a \ ,$$

und die folgenden Gesetze für 0 und 1:

$$0 \cdot a = 0 \qquad 1 + a = 1 \ .$$

Lemma 3.9

Sei $\mathcal{B}$ eine Boolesche Algebra und sei B die Trägermenge von $\mathcal{B}$. Für alle $a, b \in B$ gelten die folgenden *Absorptionsgesetze*:

$$\begin{aligned} a + ab &= a & a(a+b) &= a \\ a + {\sim}(a)b &= a + b & a({\sim}(a) + b) &= ab \end{aligned}$$

Lemma 3.10

Sei $\mathcal{B}$ eine Boolesche Algebra und sei B die Trägermenge von $\mathcal{B}$. Für alle $a, b \in B$ gilt die Eindeutigkeit des Komplements:

$$\text{Wenn } a + b = 1 \text{ und } ab = 0, \text{ dann } b = {\sim}(a) \ .$$

Beweis. Unter der Annahme von $a + b = 1$ und $ab = 0$ gelten die folgenden Gleichheiten:

$$\begin{aligned} b &= b1 = b(a + {\sim}(a)) \\ &= ba + b{\sim}(a) = 0 + b{\sim}(a) && \text{da } ab = 0 \\ &= a{\sim}(a) + b{\sim}(a) = (a + b){\sim}(a) \\ &= 1{\sim}(a) && \text{da } a + b = 1 \\ &= {\sim}(a) \ . \end{aligned}$$

□

Lemma 3.11

Sei $\mathcal{B}$ eine Boolesche Algebra und sei B die Trägermenge von $\mathcal{B}$. Für alle $a \in B$ gilt das *Involutionsgesetz*:

$${\sim}({\sim}(a)) = a \ .$$

Beweis. Nach Definition einer Booleschen Algebra ist

1. $a + {\sim}(a) = 1$ und $a \cdot {\sim}(a) = 0$ (${\sim}(a)$ ist Komplement von a)
2. ${\sim}(a) + {\sim}({\sim}(a)) = 1$ und ${\sim}(a) \cdot {\sim}({\sim}(a)) = 0$ (${\sim}({\sim}(a))$ ist Komplement von ${\sim}(a)$)

Da $+$ und $\cdot$ kommutativ folgt aus Punkt 1, dass

3. ${\sim}(a) + a = 1$ und ${\sim}(a) \cdot a = 0$ (a ist Komplement von ${\sim}(a)$)

Nun folgt aus den Punkten 2 und 3 sowie Lemma 3.10, dass ${\sim}({\sim}(a)) = a$. □

Lemma 3.12

Sei $\mathcal{B}$ eine Boolesche Algebra und sei B die Trägermenge von $\mathcal{B}$. Für alle $a, b \in B$ gelten die *Gesetze von de Morgan*:

$$ {\sim}(a + b) = {\sim}(a) \cdot {\sim}(b) \qquad {\sim}(a \cdot b) = {\sim}(a) + {\sim}(b) \; . $$

Beweis. Wir zeigen nur die erste Gleichung, die zweite überlassen wir der Leserin. Zunächst zeigen wir $(a + b) + {\sim}(a) \cdot {\sim}(b) = 1$:

$$\begin{aligned} (a + b) + {\sim}(a) \cdot {\sim}(b) &= (a + b + {\sim}(a))(a + b + {\sim}(b)) \\ &= (a + {\sim}(a) + b)(a + b + {\sim}(b)) \\ &= (1 + b)(a + 1) \\ &= 1 \cdot 1 = 1 \; . \end{aligned}$$

Nun zeigen wir $(a + b) \cdot {\sim}(a) \cdot {\sim}(b) = 0$:

$$\begin{aligned} (a + b) \cdot {\sim}(a) \cdot {\sim}(b) &= a \cdot {\sim}(a) \cdot {\sim}(b) + b \cdot {\sim}(a) \cdot {\sim}(b) \\ &= a \cdot {\sim}(a) \cdot {\sim}(b) + {\sim}(a) \cdot b \cdot {\sim}(b) \\ &= 0 \cdot {\sim}(b) + {\sim}(a) \cdot 0 \\ &= 0 + 0 = 0 \; . \end{aligned}$$

Zusammen haben wir die Voraussetzungen von Lemma 3.10 gezeigt. Somit ist ${\sim}(a) \cdot {\sim}(b)$ das Komplement von $a + b$. □

Jeder Boolesche Ausdruck F über n Boolesche Variablen repräsentiert eine Boolesche Funktion $f : \mathbb{B}^n \to \mathbb{B}$ und umgekehrt.

Definition 3.14

Sei $\mathbb{B} = \{0, 1\}$.

1. Sei F ein Boolescher Ausdruck in den Variablen $x_1, \ldots, x_n$ und bezeichne $F(s_1, \ldots, s_n)$ die Instanz von F, die wir durch Ersetzung von $x_1, \ldots, x_n$ durch $s_1, \ldots, s_n$ erhalten, wobei $s_i \in \mathbb{B}$ für alle $i = 1, \ldots, n$.

   Wir definieren die Funktion $f : \mathbb{B}^n \to \mathbb{B}$ wie folgt, wobei die Zeichen $+$, $\cdot$ und ${\sim}$ wie in Abbildung 3.1 interpretiert werden:

   $$ f(s_1, \ldots, s_n) := F(s_1, \ldots, s_n) \; . $$

   Dann heißt f die *Boolesche Funktion* zum Ausdruck F.

2. Sei $f\colon \mathbb{B}^n \to \mathbb{B}$ eine Boolesche Funktion und sei F ein Boolescher Ausdruck, dessen Boolesche Funktion gleich f. Dann nennen wir F den Booleschen Ausdruck von f.

Boolesche Funktionen erlauben uns eine direkte Definition der Äquivalenz von Booleschen Ausdrücken.

Satz 3.1

Seien F, G Boolesche Ausdrücke (in den Variablen $x_1, \ldots, x_n$) und seien $f\colon \mathbb{B}^n \to \mathbb{B}$, $g\colon \mathbb{B}^n \to \mathbb{B}$ ihre Booleschen Funktionen. Dann gilt (für eine beliebige zugrundeliegende Boolesche Algebra) $F \approx G$ gdw. $f = g$ in der Algebra der Booleschen Funktionen. □

Nach Definition sind zwei Boolesche Ausdrücke äquivalent (für die betrachtete Boolesche Algebra $\mathcal{B}$), wenn das Einsetzen von Elementen aus $\mathcal{B}$ zum gleichen Ergebnis führt. Satz 3.1 erlaubt es nun diesen Zusammenhang in Bezug auf eine spezielle Boolesche Algebra, der Algebra der Booleschen Funktionen, zu verifizieren und so Äquivalenzen für alle Booleschen Algebren zu erhalten. Grundlage des Satzes ist der Darstellungssatz von Stone.

Satz 3.2

Sei $\mathcal{B} = \langle B; +, \cdot, \sim, 0, 1\rangle$ eine Boolesche Algebra. Dann existiert eine Menge M, sodass $\mathcal{B}$ isomorph zur Mengenalgebra $\langle \mathcal{P}(M); \cup, \cap, \sim, \varnothing, M\rangle$. □

In Kapitel 2 haben wir die konjunktive und disjunktive Normalform eingeführt. Diese Begrifflichkeiten werden auch für Boolesche Algebren verwendet.

Definition 3.15

1. Ein *Literal* ist eine Boolesche Variable x oder ihre Negation $\sim(x)$
2. Ein *Summenterm* ist ein Boolescher Ausdruck der Gestalt:
$$l_1 + \cdots + l_n$$
wobei l_i Literale
3. Ein *Produktterm* ist ein Boolescher Ausdruck der Gestalt:
$$l_1 \cdot \ldots \cdot l_n$$
4. Ein Boolescher Ausdruck A ist in *konjunktiver Normalform* (*KNF*), wenn A das Produkt von Summentermen ist.
5. Ein Boolescher Ausdruck A ist in *disjunktiver Normalform* (*DNF*), wenn A die Summe von Produkttermen ist.

Wir erhalten den folgenden Satz.

Satz 3.3

Jeder Boolesche Ausdruck hat eine konjunktive beziehungsweise eine disjunktive Normalform. □

3.3. Zusammenfassung

Das Wort „Algebra“ kommt von dem arabischen Wort „al-jabr“ im Titel des Lehrbuches „Hisàb al-jabr w'al-muqâbala“, geschrieben um 820 vom Mathematiker und Astronom Al-Khowârizmi. Der Titel bedeutet etwa „Berechnungen durch Sanierung und Vereinfachung“. Wobei „Sanierung“, in Arabisch „al-jabr“, das vielfache Kürzen von Gleichungen bedeutet. Es soll nicht unerwähnt bleiben, dass das Wort „Algorithmus“ auf eine fehlerhafte Übersetzung eines anderen Lehrbuches von Al-Khowârizmi zurückgeht. Statt den Titel des Buches zu übersetzen wurde der Name des Autors als Titel angegeben.

Untersuchungen zur Booleschen Algebra gehen auf den englischen Philosophen George Boole (1815–1864) zurück, der in seinem Hauptwerk „An Investigation of the Laws of Thought“ als Erster ähnliche Strukturen untersucht hat. Wie die Logik ist auch die Boolesche Algebra ursprünglich ein mathematisches Fachgebiet, das in der Informatik Anwendung findet.

Neben denen im Anhang in Abschnitt A.1 angesprochenen Anwendungen zur Vereinfachung von logischen Schaltkreisen, findet die Boolesche Algebra Anwendung in der Logik, in der Theorie der formalen Sprachen, in der Programmierung sowie in der Statistik. Häufige Anwendungsbereiche von Algebren in der Programmierung sind etwa *abstrakte Datentypen*. Ein abstrakter Datentyp ist benutzerdefiniert und besteht aus einer Menge von Objekten, wie etwa Listen und Operationen auf diesen Objekten. Diese Datentypen werden *abstrakt* genannt, da die *Objekte* und die *Operationen* auf diesen Objekten im Vordergrund stehen. So kann ein Datentyp unabhängig von seiner Implementierung beschrieben werden.

3.4. Aufgaben

Übungsaufgabe 3.1

Betrachten Sie die Algebra $\mathcal{A} = \langle \{0, 1, 2\}; \bullet, ! \rangle$ mit

$\bullet$	0	1	2
0	0	0	1
1	0	1	2
2	0	2	0

$!$	
0	0
1	0
2	0

1. Welche der folgenden Ausdrücke sind *algebraische Ausdrücke*?
   a) x_1
   b) x_2
   c) 0
   d) !
   e) $!(x_1)$
   f) $\bullet(0, x_2)$

g) $\bullet(x_1, !(x_2))$
h) $\bullet(x_1, x_2)$
i) $\bullet(!(x_1), x_2)$

2. Welche algebraischen Ausdrücke aus Punkt 1 sind äquivalent?

Lösung.

1. a) ✓
   b) ✓
   c) ✗
   d) ✗
   e) ✓
   f) ✗
   g) ✓
   h) ✓
   i) ✓

2. Es gilt $!(x_1) \approx \bullet(x_2, !(x_3))$, da alle Instanzen der Ausdrücke den selben Wert annehmen (hier: 0).

Übungsaufgabe 3.2

Betrachten Sie die Mengenalgebra $\langle \mathcal{P}(M), \cup, \cap, \sim, \varnothing, M \rangle$. Bestimmen Sie jeweils

1. das Nullelement für die Operationen $\cup$, $\cap$ und $\sim$ sowie
2. das neutrale Element für die Operationen $\cup$, $\cap$ und $\sim$.
3. Welche algebraischen Gesetze gelten für das Mengenkomplement $\sim$ (Definition 3.9)?
4. Gibt es für die Operation $\sim$ ein Null- beziehungsweise ein neutrales Element?

Lösung.

1. Nullelement für $\cup$ ist M, da $A \cup M = M \cup A = M$. Nullelement für $\cap$ ist $\varnothing$, da $A \cap \varnothing = \varnothing \cap A = \varnothing$. Da $\sim$ keine binäre Operation ist, ist kein Nullelement definiert.

2. Einselement für $\cup$ ist $\varnothing$, da $A \cup \varnothing = \varnothing \cup A = A$. Einselement für $\cap$ ist M, da $A \cap M = M \cap A = A$. Da $\sim$ keine binäre Operation ist, ist kein Einselement definiert.

3. Da $\sim(A) := \{x \in M \mid x \notin A\}$ sind die algebraischen Gesetze $A \cup \sim(A) = M$ und $A \cap \sim(A) = \varnothing$ erfüllt.

4. Nein, da diese nur für binäre Operationen definiert sind.

Übungsaufgabe 3.3

Welche Kriterien muss eine *Boolesche Algebra* erfüllen? Beweisen Sie Lemma 2.4.
Hinweis: Zeigen Sie, dass die binäre Algebra (Definition 3.10) eine Boolesche Algebra ist. Prüfen Sie dazu alle Eigenschaften der Definition 3.7.

Lösung. Wir müssen zeigen, dass

1. $\langle \mathbb{B}; +, 0\rangle$ sowie $\langle \mathbb{B}; \cdot, 1\rangle$ kommutative Monoide sind.
2. $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ und $a + (b \cdot c) = (a + b) \cdot (a + c)$
3. für alle $a \in \mathbb{B}$ gilt $a + \sim(a) = 1$ und $a \cdot \sim(a) = 0$.

Zu a). Zuerst betrachten wir $\langle \mathbb{B}; +, 0\rangle$. Wir zeigen, dass 0 das neutrale Element bezüglich $+$ ist, d.h. für alle $a \in \mathbb{B}$ gilt $a + 0 = 0 + a = a$. Dazu betrachten wir alle möglichen Werte, die a annehmen kann und überprüfen mittels der Definition von $+$ (Abbildung 2.1.).

- Für $a = 1$
$$1 + 0 = 0 + 1 = 1 = a$$
- Für $a = 0$
$$0 + 0 = 0 + 0 = 0 = a$$

Als nächstes betrachten wir die Assoziativiät von $+$, d.h. für alle $a, b, c \in \mathbb{B}$ gilt $(a + b) + c = a + (b + c)$. Dazu unterscheiden wir die folgenden beiden Fälle:

- Für $b = 1$ müssen wir zeigen $(a + 1) + c = a + (1 + c)$:
$$(a + 1) + c = 1 + c = 1 = a + 1 = a + (1 + c)\ .$$
Hier haben wir zweimal die Definition von $+$ angewandt.
- Für $b = 0$ müssen wir zeigen $(a + 0) + c = a + (0 + c)$:
$$(a + 0) + c = a + c = a + (0 + c)\ .$$
Hier haben wir ausgenützt, dass 0 das neutrale Element von $+$ ist.

Durch den Beweis der Assoziativität von $+$ sowie der Existenz des neutralen Elements haben wir bewiesen, dass $\langle \mathbb{B}; +, 0\rangle$ ein Monoid ist. Nun müssen wir noch die Kommutativität zeigen, d.h. $a + b = b + a$ für alle $a, b \in \mathbb{B}$. Dies liest man leicht aus der Operationstafel für $+$ ab. Somit haben wir gezeigt, dass $\langle \mathbb{B}; +, 0\rangle$ ein kommutatives Monoid ist. Analog zeigt man, dass $\langle \mathbb{B}; \cdot, 1\rangle$ ein kommutatives Monoid ist. Zu b). Nun müssen wir überprüfen, dass $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ für alle $a, b, c \in \mathbb{B}$.

- Für $a = 1$ müssen wir zeigen $1 \cdot (b + c) = (1 \cdot b) + (1 \cdot c)$:
$$1 \cdot (b + c) = b + c = (1 \cdot b) + c = (1 \cdot b) + (1 \cdot c)\ .$$
Hier haben wir ausgenützt, dass 1 das neutrale Element von $\cdot$ ist.

– Für $a = 0$ müssen wir zeigen $0 \cdot (b + c) = (0 \cdot b) + (0 \cdot c)$:

$$0 \cdot (b + c) = 0 = 0 + 0 = (0 \cdot b) + (0 \cdot c) .$$

Analog zeigen wir $a + (b \cdot c) = (a + b) \cdot (a + c)$ für alle $a, b, c \in \mathbb{B}$. Zu c). Zuletzt müssen wir überprüfen, dass für alle $a \in \mathbb{B}$ gilt $a + \sim(a) = 1$ und $a \cdot \sim(a) = 0$.

– Für $a = 1$ folgt $1 + \sim(1) = 1$ nach Definition von $+$.

– Für $a = 0$ folgt $0 + \sim(0) = 1$ nach Definition von $+$.

Analog zeigt man $a \cdot \sim(a) = 0$.

Übungsaufgabe 3.4

Seien $A = \{a, b\}$ und $B = \{a, c\}$. Berechnen Sie

1. $A \cap B$
2. $A \cup B$
3. $\sim(a)$ bezüglich $\{a, b, c, d\}$
4. $A \times B$
5. $A \times \varnothing$
6. A^3

Lösung.

1. $\{a\}$
2. $\{a, b, c\}$
3. $\{c, d\}$
4. $\{(a, a), (a, c), (b, a), (b, c)\}$
5. $\varnothing$
6. $\{(a, a, a), (a, a, b), (a, b, a), (a, b, b), (b, a, a), (b, a, b), (b, b, a), (b, b, b)\}$

Übungsaufgabe 3.5

Betrachten Sie eine Boolesche Algebra $\langle B; +, \cdot, \sim, 0, 1\rangle$. Beweisen Sie zwei der vier Gesetze von Lemma 3.9.
Hinweis: Verwenden Sie die definierenden Eigenschaften einer Booleschen Algebra (Definition 3.7) um die entsprechenden Gesetze herzuleiten. Sie können auch die Gesetze von Lemma 3.8 verwenden.

Lösung.

–
$$\begin{aligned} a + ab &= a1 + ab && \langle B; \cdot, 1\rangle \text{ Monoid} \\ &= a(1 + b) && \text{Definition 2.7.2} \\ &= a1 && \text{Lemma 2.8} \\ &= a && \langle B; \cdot, 1\rangle \text{ Monoid} \end{aligned}$$

– Man kann diesen Fall analog zum ersten zeigen. Hier geben wir einen alternativen Beweis.
$$\begin{aligned} a(a + b) &= aa + ab && \text{Definition 3.7} \\ &= a + ab && \text{Lemma 3.8} \\ &= a1 + ab && \langle B; \cdot, 1\rangle \text{ Monoid} \\ &= a(1 + b) && \text{Definition 3.7} \\ &= a1 && \text{Lemma 3.8} \\ &= a && \langle B; \cdot, 1\rangle \text{ Monoid} \end{aligned}$$

–
$$\begin{aligned} a + \sim(a)b &= (a + \sim(a))(a + b) && \text{Definition 3.7} \\ &= 1(a + b) && \text{Definition 3.7} \\ &= a + b && \langle B; \cdot, 1\rangle \text{ Monoid} \end{aligned}$$

–
$$\begin{aligned} a(\sim(a) + b) &= a \sim(a) + ab && \text{Definition 3.7} \\ &= 0 + ab && \text{Definition 3.7} \\ &= ab && \langle B; +, 0\rangle \text{ Monoid} \end{aligned}$$

Übungsaufgabe 3.6

Geben Sie alle Abbildungen von $\mathbb{B}^n$ nach $\mathbb{B}^m$ an.

1. Für $n = m = 1$.
2. Für $n = 2$ und $m = 1$.

Lösung. $\mathbb{B}^2 = \{00, 01, 10, 11\}$ und $\mathbb{B}^1 = \{0, 1\}$ (wir lassen Klammern und Beistriche in der Tupelschreibweise weg). Wir erhalten

1. $f_1 \colon \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 0, 01 \mapsto 0, 10 \mapsto 0, 11 \mapsto 0$
2. $f_2 \colon \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 0, 01 \mapsto 0, 10 \mapsto 0, 11 \mapsto 1$
3. $f_3 \colon \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 0, 01 \mapsto 0, 10 \mapsto 1, 11 \mapsto 0$
4. $f_4 \colon \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 0, 01 \mapsto 0, 10 \mapsto 1, 11 \mapsto 1$
5. $f_5 \colon \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 0, 01 \mapsto 1, 10 \mapsto 0, 11 \mapsto 0$
6. $f_6 \colon \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 0, 01 \mapsto 1, 10 \mapsto 0, 11 \mapsto 1$
7. $f_7 \colon \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 0, 01 \mapsto 1, 10 \mapsto 1, 11 \mapsto 0$

8. $f_8 : \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 0, 01 \mapsto 1, 10 \mapsto 1, 11 \mapsto 1$
9. $f_9 : \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 1, 01 \mapsto 0, 10 \mapsto 0, 11 \mapsto 0$
10. $f_{10} : \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 1, 01 \mapsto 0, 10 \mapsto 0, 11 \mapsto 1$
11. $f_{11} : \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 1, 01 \mapsto 0, 10 \mapsto 1, 11 \mapsto 0$
12. $f_{12} : \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 1, 01 \mapsto 0, 10 \mapsto 1, 11 \mapsto 1$
13. $f_{13} : \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 1, 01 \mapsto 1, 10 \mapsto 0, 11 \mapsto 0$
14. $f_{14} : \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 1, 01 \mapsto 1, 10 \mapsto 0, 11 \mapsto 1$
15. $f_{15} : \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 1, 01 \mapsto 1, 10 \mapsto 1, 11 \mapsto 0$
16. $f_{16} : \mathbb{B}^2 \to \mathbb{B}^1, 00 \mapsto 1, 01 \mapsto 1, 10 \mapsto 1, 11 \mapsto 1$

Übungsaufgabe 3.7

Zeigen Sie, dass die binäre Algebra (Definition 3.10) eine Boolesche Algebra ist.

Übungsaufgabe 3.8

Zeigen Sie, dass die Algebra des n-fache kartesische Produkt von $\mathbb{B}$ (Definition 3.12) eine Boolesche Algebra ist.

Übungsaufgabe 3.9

Zeigen Sie, dass die Algebra der Abbildungen von $\mathbb{B}^n$ nach $\mathbb{B}^m$ (Definition 3.13) eine Boolesche Algebra ist.

Hinweis: Zeigen Sie, dass die Algebra des n-fache kartesische Produkt eine Boolesche Algebra ist (Aufgabe 3.8).

4.

Einführung in die Theorie der Formalen Sprachen

In diesem Kapitel werden *formale Sprachen* und *Grammatiken* eingeführt. Im Weiteren wird die *Chomsky-Hierarchie* definiert und es werden zwei Sprachklassen der Chomsky-Hierarchie, die *regulären* und die *kontextfreien* Sprachen, näher betrachtet. In Abschnitt 4.1 liegt unser Hauptaugenmerk auf den Grundbegriffen der formalen Sprachen. In Abschnitt 4.2 erklären wir, wie formale Sprachen mittels Grammatiken erzeugt werden können und typische Klassen von Sprachen werden eingeführt. In Abschnitt 4.3 konzentrieren wir uns auf eine sehr einfache Klasse von formalen Sprachen, den regulären Sprachen, und beschreiben alternative Repräsentationsformen dieser Sprachen. In Abschnitt 4.4 behandeln wir kurz kontextfreie Sprachen und besprechen eine Anwendung von diesen in Abschnitt A.2.

Schließlich stellen wir in Abschnitt 4.6 die betrachteten Konzepte in einen historischen Kontext und gehen kurz auf die Bedeutung der formalen Sprachen für die Informatik ein. Außerdem finden sich in Abschnitt 4.7 (optionale) Aufgaben zu den Themenbereichen dieses Kapitels, die zur weiteren Vertiefung dienen sollen.

4.1. Alphabete, Wörter, Sprachen

Ein *Alphabet* Σ ist eine endliche, nicht leere Menge von Symbolen (oft auch Zeichen oder Buchstaben genannt). Gemäß Konvention wird ein Alphabet durch das Symbol Σ dargestellt.

Ein *Wort* (auch *Zeichenreihe* oder *String* genannt) über Σ ist eine endliche Folge von Symbolen aus Σ. Das *Leerwort* bezeichnet das kleinste vorstellbare Wort: ein Wort ohne Buchstaben. Das Leerwort wird mit ϵ dargestellt.

Konvention

Wir verwenden üblicherweise Buchstaben vom Anfang des lateinischen Alphabets (a, b, ...), um Elemente des Alphabets zu beschreiben. Im weiteren schreiben wir Buchstaben vom Ende des Alphabets (x, y, ...), um Zeichenreihen zu bezeichnen. Um Verwechslungen auszuschließen führen wir die Konvention ein, dass $\epsilon \notin \Sigma$.

Definition 4.1

Die *Länge* eines Wortes w ist als die Anzahl der Positionen in w definiert. Die Länge von w wird mit $|w|$ bezeichnet; das Leerwort ϵ hat die Länge 0.

Definition 4.2

Wenn Σ ein Alphabet ist, können wir die Menge aller Wörter einer bestimmten Länge über Σ durch eine Potenznotation bezeichnen. Wir definieren Σ^k als die Menge der Wörter der Länge k, deren Symbole aus Σ stammen. Wir verwenden auch die folgenden Definitionen:

$$\Sigma^+ := \Sigma^1 \cup \Sigma^2 \cup \cdots$$
$$\Sigma^* := \Sigma^+ \cup \{\epsilon\}$$

Jedes Wort w über Σ ist Element von Σ^*.

Formal gilt es zwischen dem *Alphabet* Σ und Σ^1, der Menge der *Wörter* mit Länge 1 über dem Alphabet Σ, zu unterscheiden. Wir identifizieren jedoch Zeichen des Alphabets mit den Wörtern der Länge 1.

Definition 4.3

Angenommen x, y sind Wörter über Σ, dann schreiben wir $x \cdot y$ für die *Konkatenation* von x und y und definiere diese rekursiv:

$$\epsilon \cdot x = x$$
$$(ax) \cdot y = a(x \cdot y)$$

Hier bezeichnet a einen Buchstaben in Σ.
Üblicherweise schreibt man die Wörter x und y direkt hintereinander als xy, das Zeichen für die Konkatenationsoperation $\cdot$ wird also weggelassen.

Sei Σ ein Alphabet. Wir betrachten die Algebra $\langle \Sigma^*; \cdot, \epsilon \rangle$, wobei $\cdot$ wie in Definition 4.3 definiert ist. Die Konkatenation ist assoziativ und besitzt das Leerwort ϵ als neutrales Element. Also ist die Algebra $\langle \Sigma^*; \cdot, \epsilon \rangle$ ein Monoid. Diese Algebra wird auch als *Wortmonoid* bezeichnet.

Definition 4.4: Formale Sprache

Eine Teilmenge L von Σ^* heißt eine *formale Sprache* über dem *Alphabet* Σ.

Definition 4.5

Seien L, M formale Sprachen über dem Alphabet Σ. Die *Vereinigung* von L und M ist wie in der Mengenlehre definiert:

$$L \cup M := \{x \mid x \in L \text{ oder } x \in M\} .$$

Wir definieren das *Komplement von* L:

$$\sim L = \Sigma^* \setminus L := \{x \in \Sigma^* \mid x \notin L\} .$$

Der *Durchschnitt* von L und M ist wie folgt definiert:

$$L \cap M := \{x \mid x \in L \text{ und } x \in M\} .$$

Das *Produkt* von L und M, auch *Verkettung* von L und M genannt, ist definiert als:

$$LM := \{xy \mid x \in L \text{ und } y \in M\} .$$

Das nächste Lemma folgt unmittelbar aus den Definitionen und der Tatsache, dass $\langle \Sigma^*; \cdot \rangle$ ein Monoid ist.

Lemma 4.1

Seien L, L_1, L_2, L_3 formale Sprachen, dann gilt

$$(L_1L_2)L_3 = L_1(L_2L_3) \qquad L\{\epsilon\} = \{\epsilon\}L = L .$$

In der nächsten Definition erweitern wir die Potenznotation für das Alphabet Σ, siehe Definition 4.2, auf Sprachen.

Definition 4.6

Sei $L \subseteq \Sigma^*$ eine formale Sprache und $k \in \mathbb{N}$. Dann ist die *k-te Potenz* von L definiert als:

$$L^k := \begin{cases} \{\epsilon\} & \text{falls } k = 0 \\ \underbrace{LL \cdots L}_{k\text{-mal}} & \text{falls } k \geqslant 1 \end{cases}$$

Der *Kleene-Stern* * (der *Abschluss*) von L ist wie folgt definiert:

$$L^* := \bigcup_{k \geqslant 0} L^k = \{x_1 \cdots x_k \mid x_1, \ldots, x_k \in L \text{ und } k \in \mathbb{N}, k \geqslant 0\} .$$

Und wir definieren:

$$L^+ := \bigcup_{k \geqslant 1} L^k = \{x_1 \cdots x_k \mid x_1, \ldots, x_k \in L \text{ und } k \in \mathbb{N}, k \geqslant 1\} .$$

4.2. Grammatiken und Formale Sprachen

Grammatiken sind nützliche Modelle zum Entwurf von Software, die Daten mit einer rekursiven Struktur verarbeiten. Das bekannteste Beispiel ist ein *Parser*, also die Komponente eines Compilers, die mit den rekursiv verschachtelten Elementen einer typischen Programmiersprache umgeht, wie beispielsweise arithmetische Ausdrücke und Bedingungsausdrücke.

Vereinfacht ausgedrückt dienen Grammatiken als Regelwerk zur Bildung von *Sätzen* einer Sprache. Die Grammatik der deutschen Sprache etwa ist ein meist als höchst kompliziert empfundenes Regelwerk zur richtigen Bildung deutscher Sätze, die wiederum die deutsche Sprache ausmachen.

Vereinfacht können Sätze als Sequenzen von Wörtern verstanden werden. Also beschreiben Grammatiken, abstrakt gesprochen das „richtige“ Bilden von Mengen von Buchstabensequenzen, also etwa formalen Sprachen. Im Allgemeinen ist unser Interesse aber nicht auf die syntaktische Simulierung von real existierenden Sprachen bezogen, sondern vielmehr auf die Analyse rekursiver Strukturen, wie etwa Programmiersprachen, gerichtet.

Definition 4.7: Grammatik

Eine Grammatik G ist ein Quadrupel $G = (V, \Sigma, R, S)$, wobei

1. V eine endliche Menge von *Variablen* (oder *Nichtterminale*),
2. Σ ein Alphabet, die *Terminale*, $V \cap \Sigma = \varnothing$,
3. R eine endliche Menge von *Regeln*.
4. $S \in V$ das *Startsymbol* von G.

Eine Regel ist ein Paar $P \to Q$ von Wörtern, sodass $P, Q \in (V \cup \Sigma)^*$ und in P mindestens eine Variable vorkommt. Wir nennen P auch die *Prämisse* und Q die *Konklusion* der Regel.

Konvention

Variablen werden üblicherweise als Großbuchstaben geschrieben und Terminale als Kleinbuchstaben. Wenn es mehrere Regeln mit der gleichen Prämisse gibt, werden die Konklusionen auf der rechten Seite zusammengefasst: Statt $P \to Q_1$, $P \to Q_2$, $P \to Q_3$ schreiben wir kurz $P \to Q_1 \mid Q_2 \mid Q_3$.

Definition 4.8

Sei $G = (V, \Sigma, R, S)$ eine Grammatik und $x, y \in (V \cup \Sigma)^*$. Dann heißt y aus x in G *direkt ableitbar*, wenn gilt:

$$\exists\, u, v \in (V \cup \Sigma)^*,\ \exists\, (P \to Q) \in R \ \text{ sodass } \ (x = uPv \text{ und } y = uQv)\ .$$

In diesem Fall schreiben wir kurz $x \underset{G}{\Rightarrow} y$. Wenn die Grammatik G aus dem Kontext folgt, dann schreiben wir $x \Rightarrow y$.

Definition 4.9

Sei $G = (V, \Sigma, R, S)$ eine Grammatik und $x, y \in (V \cup \Sigma)^*$. Dann ist y aus x in G *ableitbar*, wenn es eine natürliche Zahl $k \in \mathbb{N}$ und Wörter $w_0, w_1, \ldots, w_k \in (V \cup \Sigma)^*$ gibt, sodass

$$x = w_0 \underset{G}{\Rightarrow} w_1 \underset{G}{\Rightarrow} \ldots \underset{G}{\Rightarrow} w_k = y\ ,$$

das heißt $x = y$ für $k = 0$. Symbolisch schreiben wir $x \overset{*}{\underset{G}{\Rightarrow}} y$, beziehungsweise $x \overset{*}{\Rightarrow} y$.

Die vom Startsymbol S ableitbaren Wörter heißen *Satzformen*. Elemente von Σ^* werden *Terminalwörter* genannt. Satzformen, die Terminalwörter sind, heißen *Sätze*. Sätze können mehrere Ableitungen haben und es kann Satzformen geben, die nicht weiter abgeleitet werden können.

Definition 4.10: Sprache einer Grammatik

Die Menge aller Sätze

$$\mathsf{L}(G) := \{x \in \Sigma^* \mid S \overset{*}{\underset{G}{\Rightarrow}} x\} ,$$

wird die von der Grammatik G *erzeugte Sprache* genannt. Zwei Grammatiken G_1 und G_2 heißen *äquivalent*, wenn $\mathsf{L}(G_1) = \mathsf{L}(G_2)$ gilt.

Anhand der zugelassenen Form der Regeln unterscheidet man verschiedene Klassen von Grammatiken.

Definition 4.11

Sei $G = (V, \Sigma, R, S)$ eine Grammatik. Dann heißt G

- *rechtslinear*, wenn für alle Regeln $P \to Q$ in R gilt, dass $P \in V$ und $Q \in \Sigma^* \cup \Sigma^+ V$,
- *kontextfrei*, wenn für alle Regeln $P \to Q$ gilt, dass $P \in V$ und $Q \in (V \cup \Sigma)^*$,
- *kontextsensitiv*, wenn für alle Regeln $P \to Q$ gilt:
  1. entweder es existieren $u, v, w \in (V \cup \Sigma)^*$ und $A \in V$, sodass
  $$P = uAv \text{ und } Q = uwv \text{ wobei } |w| \geqslant 1 ,$$
  2. oder $P = S$ und $Q = \epsilon$,

  Wenn $S \to \epsilon \in R$, dann kommt S nicht in einer Konklusion vor.
- *beschränkt*, wenn für alle Regeln $P \to Q$ entweder gilt:
  1. $|P| \leqslant |Q|$ oder
  2. $P = S$ und $Q = \epsilon$.

  Wenn $S \to \epsilon \in G$, dann kommt S nicht in einer Konklusion vor.

Aufbauend auf die eingeführten Klassen von Grammatiken, werden entsprechend Klassen von formalen Sprachen definiert.

Definition 4.12

Eine formale Sprache L heißt

- *regulär* oder vom *Typ 3*, wenn eine rechtslineare Grammatik G existiert, sodass $L = \mathsf{L}(G)$,
- *kontextfrei* oder vom *Typ 2*, wenn eine kontextfreie Grammatik G existiert, sodass $L = \mathsf{L}(G)$,
- *kontextsensitiv* oder vom *Typ 1*, wenn eine kontextsensitive Grammatik G existiert, sodass $L = \mathsf{L}(G)$,
- *beschränkt*, wenn eine beschränkte Grammatik G existiert, sodass $L = \mathsf{L}(G)$,

– *rekursiv aufzählbar* oder vom *Typ 0*, wenn eine Grammatik G existiert, sodass $L = \mathsf{L}(G)$.

Zu beachten ist, dass es formale Sprachen gibt, die gar nicht durch eine Grammatik beschrieben werden können. Für formale Sprachen gelten die folgenden Inklusionen, die als *Chomsky-Hierarchie* bekannt sind.

$$\mathcal{L}_3 \subsetneq \mathcal{L}_2 \subsetneq \mathcal{L}_1 \subsetneq \mathcal{L}_0 \subsetneq \mathcal{L} \,,$$

wobei $\mathcal{L}_i$ $(i = 0, 1, 2, 3)$ die Klasse der Sprachen von Typ i und $\mathcal{L}$ die Klasse der formalen Sprachen bezeichnet.

Satz 4.1

Die Chomsky-Hierarchie ist eine Hierarchie, das heißt alle Inklusionen gelten und sind strikt.

Beweis. Wir werden hier nur den Beweis führen, dass alle Inklusionen gelten. Für die entscheidende und wesentlich aufwändigere Argumentation, dass alle Inklusionen strikt sind, verweisen wir auf [6].

Anhand der Definitionen der Grammatiken ist leicht einzusehen, dass eine rechtslineare Grammatik auch kontextfrei ist. Somit gilt der Zusammenhang $\mathcal{L}_3 \subseteq \mathcal{L}_2$. Jede kontextfreie Grammatik, die keine Regeln der Form $A \to \epsilon$ verwendet, ist laut Definition auch eine kontextsensitiv Grammatik. Regeln der Form $A \to \epsilon$ werden ϵ-*Regeln* (oder ϵ-*Produktionen*) genannt. Man kann zeigen, dass man jede kontextfreie Grammatik mit ϵ-Regeln in eine kontextfreie Grammatik, die auch eine kontextsensitive Grammatik ist, umschreiben kann [6]. Zusammenfassend gilt $\mathcal{L}_2 \subseteq \mathcal{L}_1$. Im weiteren gilt offensichtlich, dass eine kontextsensitive Grammatik überhaupt eine Grammatik ist, somit folgt $\mathcal{L}_1 \subseteq \mathcal{L}_0$. Schließlich beschreibt jede Grammatik eine formale Sprache, also folgt $\mathcal{L}_0 \subseteq \mathcal{L}$. □

Die Chomsky-Hierarchie erwähnt beschränkte Grammatiken nicht. Das erklärt sich durch den nächsten Satz, für dessen Beweis wir ebenfalls auf [6] verweisen.

Satz 4.2

Eine Sprache L ist kontextsensitiv gdw. L beschränkt ist. □

4.3. Reguläre Sprachen

Wir haben oben festgelegt, dass eine Sprache *regulär* heißt, wenn sie von einer rechtslinearen Grammatik beschrieben wird. Reguläre Sprachen sind die einfachste Klasse von Sprachen in der Chomsky-Hierarchie. Reguläre Sprachen haben eine derart große Bedeutung, dass neben der Charakterisierung von regulären Sprachen durch Grammatiken auch Beschreibungen mit Hilfe von *endlichen Automaten* und *regulären Ausdrücken* untersucht werden [9].

Reguläre Sprachen finden etwa in den folgenden Bereichen ihre Anwendung.

– Software zum Entwurf und Testen von *digitalen Schaltkreisen*.

– Softwarebausteine eines Compilers. Der *lexikalische Scanner* („*Lexer*“) eines Compilers wird üblicherweise mit Hilfe von endlichen Automaten implementiert und dient zur Aufteilung des Eingabetextes in logische Einheiten, wie Bezeichner oder Schlüsselwörter.

- Software zum *Durchsuchen* umfangreicher Texte, wie Sammlungen von Webseiten, um Vorkommen von Wörtern, Ausdrücken oder anderer Muster zu finden.
- Software zur Verifizierung aller Arten von Systemen, die eine endliche Anzahl verschiedener Zustände besitzen, wie Kommunikationsprotokolle oder *Protokolle* zum sicheren Datenaustausch.
- Softwarebausteine eines Computerspiels. Die Logik bei der *Kontrolle von Spielfiguren* kann mit Hilfe eines endlichen Automaten implementiert werden. Dies erlaubt eine bessere Modularisierung des Codes.

In diesem Abschnitt werden zunächst endliche Automaten informell anhand einer kleinen Anwendung eingeführt. Endliche Automaten stellen ein einfaches formales Modell dar, das trotz oder gerade wegen seiner Einfachheit vielfache Verwendung findet. Anschließend an die Motivation wird der Zusammenhang zwischen endlichen Automaten und rechtslinearen Grammatiken skizziert. Vertiefungen der präsentierten Begriffe sowie Beweise für die aufgestellten Behauptungen werden in der Lehrveranstaltung Diskrete Strukturen behandelt.

Im einführenden Beispiel untersuchen wir Protokolle, die den Gebrauch elektronischen „Geldes“ ermöglichen. Mit elektronischem „Geld“ sind Dateien gemeint, mit denen Kunden Waren im Internet bezahlen können. Es handeln drei Parteien: der *Kunde*, die *Bank* und das *Geschäft*. Die Interaktion zwischen diesen Parteien ist auf die folgenden fünf Aktionen beschränkt:

- Der Kunde kann *zahlen*, das heißt der Kunde sendet das Geld beziehungsweise weist die Bank an, an seiner Stelle zu zahlen.
- Der Kunde kann die Geldanweisung *stornieren*.
- Das Geschäft kann dem Kunden Waren *liefern*.
- Das Geschäft kann Geld *einlösen*.
- Die Bank kann Geld *überweisen*.

Wir treffen die folgenden *Grundannahmen*:

- Der Kunde ist *unverantwortlich*.
- Das Geschäft ist *verantwortlich*, aber *gutgläubig*.
- Die Bank ist *strikt*.

Die Handlungen werden in einem *Protokoll* zusammengefasst. Protokolle dieser Art lassen sich durch endliche Automaten darstellen.

- Jeder Zustand repräsentiert die *Situation* eines Partners.
- Zustandsübergänge entsprechen *Aktionen* oder *Handlungen* der Partner.
- Wir betrachten diese Handlungen als „extern“; die Sequenz der Handlungen ist wichtig, nicht wer sie initiiert.

Die Abbildungen 4.1–4.3 präsentieren die endlichen Automaten, die die Aktionen der drei Partner beschreiben. Wie in den Abbildungen bezeichnen wir diese Automaten mit G (für den Geschäftsautomaten), K (für den Kundenautomaten) und B (für den Bankautomaten).

Unsere Spezifikation des Protokolls mit endlichen Automaten ist leicht unterrepräsentiert. Etwa ist derzeit ungeklärt wie die drei Automaten zusammenwirken: In welchen Zustand soll G wechseln, wenn der Kunde beschließt die Transaktion zu stornieren? Da das Geschäft von dieser

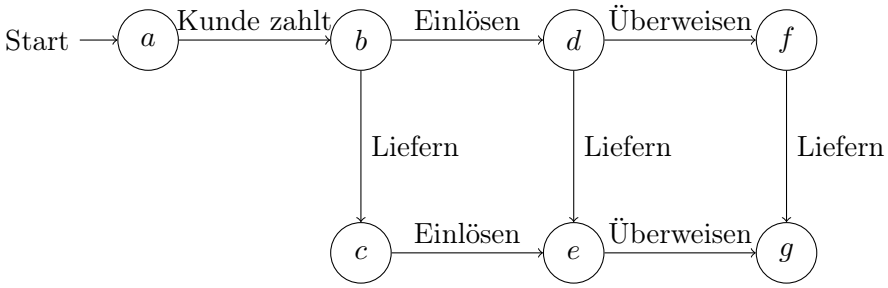


Abbildung 4.1.: Der Geschäftsautomat G.

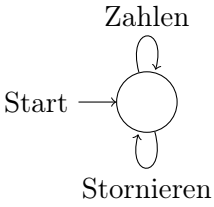


Abbildung 4.2.: Der Kundenautomat K.

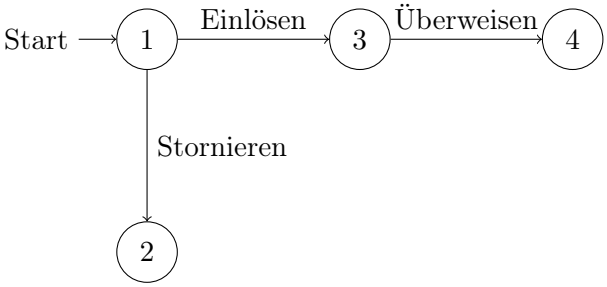


Abbildung 4.3.: Der Bankautomat B.

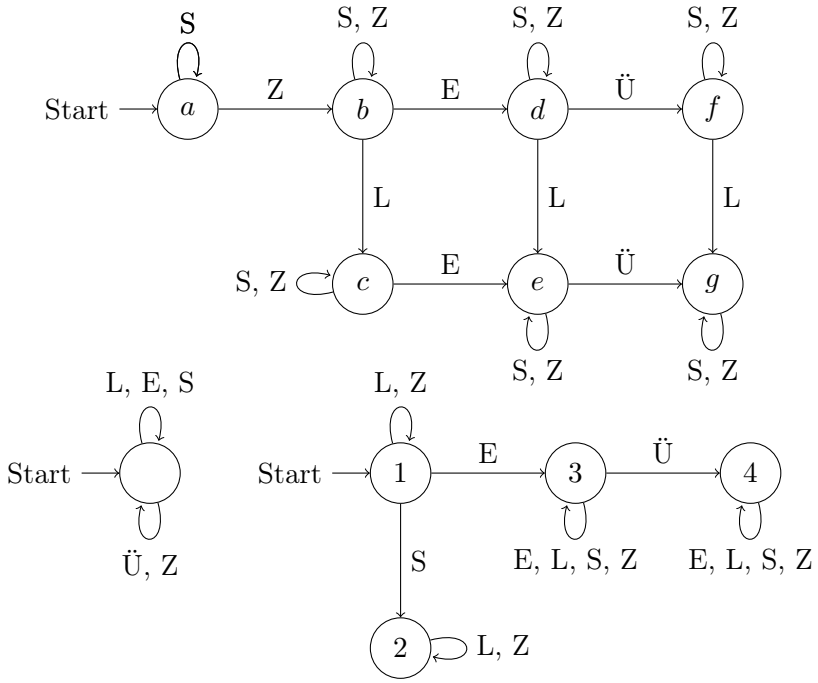


Abbildung 4.4.: Endliche Automaten G, K und B.

Aktion nicht (direkt) betroffen ist, sollte der Zustand beibehalten werden. Dies muss jedoch noch explizit vereinbart werden. Außerdem fehlen noch Übergänge für Verhalten der Agenten, die nicht beabsichtigt sind. Auch in diesem Fall muss vereinbart werden, dass die jeweiligen Automaten in ihrem bisherigen Zustand verharren. Das heißt wir müssen die Automaten explizit dazu befähigen gewisse Aktionen zu ignorieren. Die erweiterten Automaten sind in Abbildung 4.4 dargestellt, dabei verwenden wir die folgenden Abkürzungen:

Zahlen... Z Einlösen... E Stornieren... S Liefern... L Überweisen... U

In der exakten Formalisierung von endlichen Automaten (siehe Definition 4.13) wird das Problem dadurch gelöst, dass die Automaten gemeinsam auf *alle* möglichen Aktionen reagieren können müssen. Diese Lösung unterlassen wir hier, da es den erweiterten Automaten zu unübersichtlich machen würde.

Nun kombinieren wir unsere Automaten, sodass wir die Interaktionen zwischen den Agenten abbilden können. Um diese Interaktion zu modellieren, genügt es, die Interaktion zwischen dem Automaten B, der die Bankgeschäfte beschreibt, und dem Automaten G, der die Aktionen des Geschäftes darstellt, zu beschreiben. Dies geschieht, indem wir den *Produktautomaten* $B \times G$ aus B und G definieren.

1. Die *Zustände* dieses Automaten werden als Paare

$$(i, x) \colon i \in \{1, 2, 3, 4\}, x \in \{a, b, c, d, e, f, g\} \ ,$$

   angegeben.

2. Die *Übergänge* werden durch *paralleles* Ausführen von B und G definiert. Angenommen $B \times G$ ist im Zustand (i, x), das heißt B ist im Zustand i und G im Zustand x. Sei X

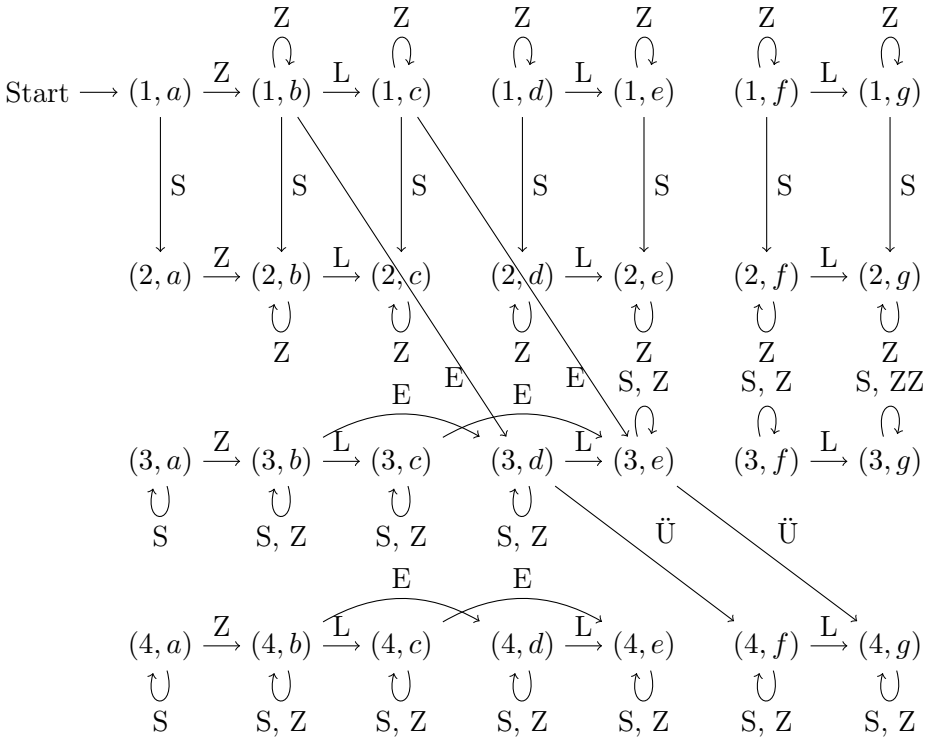


Abbildung 4.5.: Produktautomat $B \times G$.

eine Eingabe. Gelte nun, dass B mit Eingabe X in den Zustand i' wechselt. Andererseits wechselt G bei Eingabe X in den Zustand x'. Dann geht der Zustand (i, x) in $B \times G$ zu (i', x') über. Dieser Übergang wird durch eine Kante mit der Markierung X repräsentiert.

Der Automat, den wir erhalten, ist in Abbildung 4.5 dargestellt. Um Platz zu sparen, ist die Darstellung leicht vereinfacht; auf die Kreise rund um die Zustände wurde verzichtet.

Der Produktautomat $B \times G$ lässt nun einige Schlussfolgerungen zu. Einerseits ist das Protokoll ineffizient. Von den 28 möglichen Zuständen, sind nur 10 tatsächlich vom Startzustand aus erreichbar, die anderen 18 sind *unerreichbar* und können weggelassen werden.

Kritischer ist, dass das Protokoll nicht sicher ist. Der Automat $B \times G$ kann in einen Zustand gelangen, in welchem die Waren geschickt wurden und trotzdem nie eine Überweisung an das Geschäft erfolgen wird. Betrachte etwa den Zustand $(2, c)$. In diesem Zustand hat die Bank einen Antrag das elektronische Geld zu löschen (S) bearbeitet. Dies geschah, bevor die Anweisung zu überweisen (Ü) bearbeitet werden konnte. Trotzdem hat das Geschäft die Waren bereits versandt.

In diesem, sehr vereinfachendem Beispiel, können wir sehen wie endliche Automaten zur Formalisierung eines endlichen Systems verwendet werden können. Dies erlaubt es uns dann Aussagen über die Korrektheit des Systems treffen zu können.

Nach dieser informellen Einführung in die Verwendung von endlichen Automaten wenden wir uns der formalen Definition zu.

Definition 4.13: Deterministischer endlicher Automat

Ein *deterministischer endlicher Automat* (*DEA*) ist ein Quintupel $A = (Q, \Sigma, \delta, q_0, F)$, sodass

	$a_1 \in \Sigma$	$a_2 \in \Sigma$	$\cdots$
$q_1 \in Q$	$\delta(q_1, a_1)$	$\delta(q_1, a_2)$	$\cdots$
$q_2 \in Q$	$\delta(q_2, a_1)$		
$\vdots$	$\vdots$		

Abbildung 4.6.: Die Übergangsfunktion dargestellt durch die Zustandstabelle.

1. Q eine endliche Menge von *Zuständen*,
2. Σ eine endliche Menge von *Eingabesymbolen*, (Σ wird auch *Eingabealphabet* genannt)
3. $\delta\colon Q \times \Sigma \to Q$ die *Übergangsfunktion*,
4. $s \in Q$ der *Startzustand* und
5. $F \subseteq Q$ eine endliche Menge von *akzeptierenden Zuständen*.

Die Übergangsfunktion gibt an wie sich der Zustand des Automaten bei einer Eingabe ändert. Zu beachten ist, dass δ für alle möglichen Argumente definiert sein muss.

Die Übergangsfunktion kann tabellarisch in der *Zustandstabelle* angegeben werden, siehe Abbildung 4.6. Ähnlich wie in der Motivation kann der Automat durch seinen *Zustandsgraphen* visualisiert werden.

Definition 4.14

Sei $A = (Q, \Sigma, \delta, s, F)$ ein DEA, der *Zustandsgraph* ist ein gerichteter Graph, sodass

1. die Ecken die Zustände sind,
2. für Zustände $p, q \in Q$ sind die Kanten von p nach q alle Tripel

$$(p, a, q) \qquad \text{mit} \qquad a \in \Sigma \qquad \text{und} \qquad \delta(p, a) = q\ .$$

Konvention

Üblicherweise schreibt man zu jeder Kante (p, a, q) die Eingabe a, den Startzustand markiert man mit einem Pfeil und die akzeptierenden Zustände werden mit einem doppelten Kreis gekennzeichnet.

Aus Definition 4.13 ist der Zusammenhang von endlichen Automaten zu formalen Sprachen ersichtlich. Die Aktionen, die wir in einem Automaten durchführen können, werden durch Buchstaben eines Alphabets repräsentiert. Sei $A = (Q, \Sigma, \delta, q_0, F)$ und gelte $\delta(p, a) = q$, dann sagen wir der Automat A *liest* den Buchstaben a, wenn er vom Zustand p nach q wechselt. Wir erweitern diesen Zusammenhang auf das Lesen von Wörtern.

Definition 4.15

Sei $A = (Q, \Sigma, \delta, s, F)$ ein DEA. Wir definieren die *erweiterte Übergangsfunktion* $\hat{\delta}\colon Q \times \Sigma^* \to Q$ induktiv.

$$\begin{aligned} \hat{\delta}(q, \epsilon) &:= q \\ \hat{\delta}(q, xa) &:= \delta(\hat{\delta}(q, x), a) \qquad\qquad x \in \Sigma^*,\ a \in \Sigma \end{aligned}$$

Beachten Sie, dass wir hier von unserer Konvention Gebrauch machen, dass Buchstaben vom Ende des Alphabets (in der Definition etwa x) Strings bezeichnen, wohingegen lateinische Buchstaben vom Anfang des Alphabets (etwa a) einzelne Buchstaben im Alphabet bezeichnen.

Sei A wie oben definiert und gelte nun $\hat{\delta}(p, x) = q$, dann sagen wir der Automat A *liest* das Wort x am Weg von Zustand p nach Zustand q. Schließlich können wir die *Sprache* eines DEA definieren.

Definition 4.16: Sprache eines DEA

Sei Σ ein Alphabet und sei $A = (Q, \Sigma, \delta, s, F)$ ein DEA. Die Sprache $\mathsf{L}(A)$ von A ist wie folgt definiert:

$$\mathsf{L}(A) := \{x \in \Sigma^* \mid \hat{\delta}(s, x) \in F\}\ .$$

Das heißt, die Sprache von A ist die Menge der Zeichenreihen x, die vom Startzustand s in einen akzeptierenden Zustand führen. Wir nennen $\mathsf{L}(A)$ auch die von A *akzeptierte* Sprache.

Der nächste Satz stellt den Zusammenhang zwischen regulären Sprachen und Sprachen, die von einem endlichen Automaten akzeptiert werden können dar. Für den Beweis sei auf [6] verwiesen.

Satz 4.3

Für jeden DEA A ist $\mathsf{L}(A)$ regulär. Umgekehrt existiert zu jeder regulären Sprache L ein DEA A, sodass $L = \mathsf{L}(A)$. □

Für reguläre Sprachen gelten nützliche Abschlusseigenschaften.

Satz 4.4

Die Vereinigung $L \cup M$ zweier regulärer Sprachen L, M ist regulär. □

Beachten Sie, dass die Sprachen L und M in dem Satz nicht notwendigerweise über dasselbe Alphabet definiert sind. Dies stellt eine harmlose Verallgemeinerung dar, da immer die Vereinigung der jeweiligen Alphabete betrachtet werden kann.

Satz 4.5

Sei L eine reguläre Sprache über dem Alphabet Σ. Dann ist das Komplement $\sim L = \Sigma^* \setminus L$ ebenfalls regulär. □

Satz 4.6

Wenn L und M reguläre Sprachen sind, dann ist auch $L \cap M$ regulär.

Beweis. Wir können das Gesetz von de Morgan anwenden.

$$L \cap M = \sim\sim L \cup \sim M \ .$$

Mit den den Sätzen 4.4 und 4.5 folgt, dass reguläre Sprachen unter Vereinigung und Komplement abgeschlossen sind. □

Satz 4.7

Wenn L und M reguläre Sprachen sind, dann ist auch $L \setminus M$ regulär.

Beweis. Dazu verwenden wir:

$$L \setminus M = L \cap \sim M \ .$$

□

Die oben dargestellten Abschlusseigenschaften regulärer Sprachen beziehen sich alle auf Boolesche Operationen. Es gelten aber noch weitere Abschlusseigenschaften, etwa spezialisiert die folgende Definition den Homomorphismusbegriff auf formale Sprachen.

Definition 4.17: Homomorphismus

Seien Σ und Γ Alphabete. Ein *Stringhomomorphismus* ist eine Abbildung $\varphi\colon \Sigma^* \to \Gamma^*$, sodass für alle $x, y \in \Sigma^*$ gilt:

$$\varphi(xy) = \varphi(x) \cdot \varphi(y) \ .$$

Das heißt, die Abbildung φ erfüllt die Homomorphiebedingung auf der Konkatenation.

Für einen Beweis des folgenden Satzes sei auf [11] verwiesen.

Satz 4.8

Sei L eine reguläre Sprache über dem Alphabet Σ und sei $\varphi\colon \Sigma^* \to \Gamma^*$ ein Homomorphismus. Dann ist $\varphi(L)$ regulär. □

4.4. Kontextfreie Sprachen

Eine Sprache heißt *kontextfrei* wenn sie von einer kontextfreien Grammatik (kurz *KFG*) beschrieben wird. Wir skizzieren die Ausdrucksfähigkeit von kontextfreien Sprachen anhand der Sprache der Palindrome über dem Alphabet $\Sigma = \{0, 1\}$. Diese Sprache ist nicht regulär.

Definition 4.18

Induktive Definition von Palindromen über Σ:

1. ϵ, 0, 1 sind Palindrome.

2. Wenn x ein Palindrom ist, dann sind auch $\mathsf{0}x\mathsf{0}$ und $\mathsf{1}x\mathsf{1}$ Palindrome.

Die KFG $G = (\{P\}, \Sigma, R, P)$, wobei R wie folgt definiert ist, beschreibt die Sprache der Palindrome.

$$P \to \epsilon \mid \mathsf{0} \mid \mathsf{1}$$
$$P \to \mathsf{0}P\mathsf{0} \mid \mathsf{1}P\mathsf{1}$$

Satz 4.9

$\mathsf{L}(G)$ ist genau die Menge der Palindrome über dem Alphabet $\{\mathsf{0}, \mathsf{1}\}$.

Beweis. Die Behauptung ist, dass $x \in \mathsf{L}(G)$ gdw. x ein Palindrom ist. Die Richtung von links nach rechts überlassen wir der Leserin. Wir betrachten die Richtung von rechts nach links, also „Wenn x ist ein Palindrom ist, dann ist $x \in \mathsf{L}(G)$“. Der Beweis ist mittels Induktion nach $|x|$.[1]

1. Basis: Wir zeigen die Behauptung für $|x| = 0$ und $|x| = 1$ und betrachten die Worte ϵ, $\mathsf{0}$ und $\mathsf{1}$. Diese sind in $\mathsf{L}(G)$, da die Regeln

$$P \to \epsilon \qquad P \to \mathsf{0} \qquad P \to \mathsf{1}\,,$$

in R sind.

2. Schritt: Wir können $|x| \geqslant 2$ annehmen. Da x ein Palindrom ist, muss ein $y \in \Sigma^*$ existieren, sodass:

$$x = \mathsf{0}y\mathsf{0} \quad \text{oder} \quad x = \mathsf{1}y\mathsf{1}\,.$$

OBdA. sei $x = \mathsf{0}y\mathsf{0}$. Dann ist die Induktionshypothese auf y anwendbar und wir wissen, dass $y \in \mathsf{L}(G)$. Somit gilt $P \overset{*}{\Rightarrow} y$ und daher auch:

$$P \Rightarrow \mathsf{0}P\mathsf{0} \overset{*}{\Rightarrow} \mathsf{0}y\mathsf{0} = x\,.$$

□

Im Induktionschritt des obigen Beweises haben wir implizit den folgenden Sachverhalt zu Ableitungen in der Grammatik G verwendet. Wir bezeichnen diesen Vorgang als das *Einbetten* von Ableitungen.

Lemma 4.2

Sei G eine Grammatik und sei $A \overset{*}{\Rightarrow} x$ eine Ableitung in G und seien $u, v \in (V \cup \Sigma)^*$. Dann ist auch $uAv \overset{*}{\Rightarrow} uxv$ eine Ableitung in G.

Beweis. Wenn $A \overset{*}{\underset{G}{\Rightarrow}} x$, dann gibt es ein $k \in \mathbb{N}$ und $w_0, \ldots, w_k \in (V \cup \Sigma)^*$, sodass

$$A \Rightarrow w_1 \Rightarrow \cdots \Rightarrow w_{k-1} \Rightarrow x\,.$$

[1] Zum besseren Verständnis des Beweises empfiehlt es sich das Kapitel zum induktiven Beweisen in Kapitel 1.

Wir zeigen das Lemma mittels Induktion nach k.

1. Basis. Wenn $k = 0$ dann ist nichts zu zeigen.
2. Schritt. Sei $k > 0$. Wir betrachten die Ableitung $A \overset{*}{\Rightarrow} w_{k-1}$. Nach Induktionshypothese existiert eine Ableitung $uAv \overset{*}{\Rightarrow} uw_{k-1}v$. Es genügt also die Ableitung $w_{k-1} \Rightarrow x$ in die Ableitung $uw_{k-1}v \Rightarrow uxv$ umzuschreiben.

□

Sei $G = (V, \Sigma, R, S)$ eine KFG. Bei der *Linksableitung* wird in jeder Satzform das am weitesten links stehende Nichtterminalsymbol ersetzt, bei der *Rechtsableitung* das am weitesten rechts stehende Nichtterminalsymbol. Für Wörter $x, y \in (V \cup \Sigma)^*$ schreiben wir $x \underset{\ell}{\Rightarrow} y$, wenn y aus x in G gemäß einer Linksableitung direkt ableitbar ist und $x \underset{\mathrm{r}}{\Rightarrow} y$, wenn y aus x gemäß einer Rechtsableitung direkt ableitbar ist.

Definition 4.19

Eine Grammatik G heißt *eindeutig*, wenn jedes Wort $x \in \mathsf{L}(G)$ genau eine Linksableitung besitzt, ansonsten nennt man G *mehrdeutig*.

Statt Regeln von links nach rechts, also *top-down*, auszuwerten, können wir das auch von rechts nach links, also *bottom-up*, tun.

Definition 4.20

Sei $G = (V, \Sigma, R, S)$ eine KFG und sei $A \to X_1 \dots X_n$ mit $X_i \in (V \cup \Sigma)$ eine Regel in G. Die *rekursive Inferenz* $\mathsf{L}(A)$ ist induktiv definiert: Wenn $x_i \in \mathsf{L}(X_i)$ oder $X_i = x_i \in \Sigma$, dann gilt $x_1 x_2 \cdots x_n \in \mathsf{L}(A)$.

Definition 4.21: Syntaxbaum

Sei $G = (V, \Sigma, R, S)$ eine KFG. Ein *Syntaxbaum* für G ist ein Baum B, sodass die folgenden Bedingungen gelten:

1. Jeder innere Knoten von B ist eine Variable in V.
2. Jedes Blatt in B ist entweder ein Terminal aus Σ, ein Nichtterminal aus V oder ϵ. Wenn das Blatt ϵ ist, dann ist dieser Knoten das einzige Kind seines Vorgängers.
3. Sei A ein innerer Knoten, $X_1, \dots, X_n$ seine Kinder. Dann ist

$$A \to X_1 \cdots X_n \in R\,.$$

Definition 4.22

Sei $G = (V, \Sigma, R, S)$ eine KFG. Das *Ergebnis* eines Syntaxbaums B für G ist das Wort über $(V \cup \Sigma)^*$, das wir erhalten, wenn wir die Blätter in S von links nach rechts lesen.

Satz 4.10

Sei $G = (V, \Sigma, R, S)$ eine KFG, $A \in V$ und $x \in \Sigma^*$. Die folgenden Beschreibungen kontextfreier Sprachen sind *äquivalent*.

1. $x \in \mathsf{L}(A)$ nach dem rekursiven Inferenzverfahren.
2. $A \overset{*}{\Rightarrow} x$.
3. $A \overset{*}{\underset{\ell}{\Rightarrow}} x$.
4. $A \overset{*}{\underset{\mathrm{r}}{\Rightarrow}} x$.
5. Es existiert ein Syntaxbaum mit Wurzel A und Ergebnis x.

Beweis. Der Satz folgt aus den folgenden Sätzen 4.11 – 4.14. □

Im restlichen Abschnitt werden wir Satz 4.10 beweisen.

Satz 4.11

Sei $G = (V, \Sigma, R, S)$ eine KFG. Sei A ein Nichtterminal in V. Angenommen $x \in \mathsf{L}(A)$ nach dem rekursiven Inferenzverfahren, dann gibt es einen Syntaxbaum mit Wurzel A und Ergebnis x.

Beweis. Wir zeigen den Satz mit Induktion über die Anzahl der Schritte in der rekursiven Inferenz von $x \in \mathsf{L}(A)$.

1. Basis: Angenommen es war ein Schritt nötig, um $x \in \mathsf{L}(A)$ festzustellen. Dann muss es eine Regel $A \to x$ in R geben und es ist leicht einen Syntaxbaum von A zu konstruieren, dessen Ergebnis x ist.
2. Schritt: Angenommen $n + 1$ Inferenzschritte wurden durchgeführt, um $x \in \mathsf{L}(A)$ nachzuweisen. Der $n + 1^{\text{te}}$ Schritt erfordert die Existenz einer Regel $A \to X_1 X_2 \cdots X_n$ in R, wobei gelten muss: $x = x_1 x_2 \cdots x_n$ und für alle $i = 1, \ldots, n$: $x_i \in \mathsf{L}(X_i)$. Es gibt zwei Möglichkeiten:
   - Wenn X_i ein Terminalsymbol ist, dann gilt $x_i = X_i$.
   - Wenn X_i eine Variable ist, dann existiert nach Induktionshypothese ein Syntaxbaum mit Ergebnis x_i, dessen Wurzel X_i ist.

   In beiden Fällen ist leicht einzusehen, wie der Syntaxbaum für A definiert werden muss, sodass x das Ergebnis dieses Syntaxbaumes ist.

□

Satz 4.12

Sei $G = (V, \Sigma, R, S)$ eine KFG, sodass ein Syntaxbaum B mit Wurzel A und Ergebnis $x \in \Sigma^*$ existiert, dann gibt es eine Linksableitung von x aus A in G.

Beweis. Wir zeigen den Satz mittels Induktion über die Höhe des Syntaxbaumes, also der maximalen Anzahl von Kanten von der Wurzel zu einem Blatt.

1. Basis: Hat der Baum S die Höhe 1, dann gibt es jeweils nur eine Kante von der Wurzel zu den Blättern. Nach Definition 4.21 ist das nur möglich, wenn auch eine Regel $A \to x$ in R vorkommt.

2. Schritt: Angenommen S hat Höhe $n+1$. Dann existiert eine Regel $A \to X_1 X_2 \cdots X_n$ in R und es existieren Worte $x_i \in (V \cup \Sigma)^*$, sodass $x = x_1 \ldots x_n$. (Hier nehmen wir wieder an $X_i \in V \cup \Sigma$.) Im Weiteren hat S n direkte Teilbäume S_i, deren Wurzeln mit X_i markiert sind und deren Ergebnisse jeweils x_i ist. Es gelten die folgenden beiden Fälle:
   - Wenn $X_i \in \Sigma$, dann $x_i = X_i$ oder
   - $X_i \in V$, dann ist die Induktionhypothese anwendbar, und es existiert eine Linksableitung $X_i \overset{*}{\underset{\ell}{\Rightarrow}} x_i$.

   Wir konstruieren eine Linksableitung von $x = x_1 x_2 \cdots x_n$:

$$\begin{aligned} A \underset{\ell}{\Rightarrow} X_1 X_2 \cdots X_n &\overset{*}{\underset{\ell}{\Rightarrow}} x_1 X_2 \cdots X_n \\ &\vdots \\ &\overset{*}{\underset{\ell}{\Rightarrow}} x_1 x_2 \cdots x_{n-1} x_n \ . \end{aligned}$$

   Formal zeigen wir, dass für alle $i = 1, \ldots, n$ gilt

$$A \overset{*}{\underset{\ell}{\Rightarrow}} x_1 x_2 \cdots x_i X_{i+1} \cdots X_n \ .$$

□

Satz 4.13

Sei $G = (V, \Sigma, R, S)$ eine KFG, sodass ein Syntaxbaum mit Wurzel A und Ergebnis $x \in \Sigma^*$ existiert, dann gibt es eine Rechtsableitung von w aus A in G.

Beweis. Der Beweis verläuft analog zum Beweis von Satz 4.12. □

Als Vorbereitung für den nächsten Satz stellen wir fest, dass wir Ableitungen *aufbrechen* können. Aufbrechen von Ableitungen ist die zur Einbettung inverse Operation, siehe Lemma 4.2.

Lemma 4.3

Sei G eine KFG und sei $A \Rightarrow X_1 X_2 \ldots X_n \overset{*}{\Rightarrow} x$ eine Ableitung in G, wobei $X_i \in V \cup \Sigma$. Dann können wir x in die Stücke x_1, x_2, …, x_n brechen, sodass für alle $i = 1, \ldots, n$, $X_i \overset{*}{\Rightarrow} x_i$ Ableitungen in G sind.

Beweis. Für $X_i \in \Sigma$ ist die Aussage klar: Wenn X_i ein Terminalsymbol, dann $x_i = X_i$ und die Ableitung enthält keine Schritte. Sonst zeigt man zunächst (mit Induktion nach den Ableitungsschritten), dass wenn

$$X_1 X_2 \ldots X_n \overset{*}{\Rightarrow} x \ ,$$

alle Satzformen, die aus der Ersetzung von X_i in x links von Satzformen die aus der Ersetzung von X_j entstehen, wenn $i < j$. Somit, wenn $X_i \in V$, dann erhalten wir $X_i \overset{*}{\Rightarrow} x_i$, indem

– alle Positionen der Satzformen links und rechts von Positionen, die aus X_i abgeleitet werden, eliminiert werden und

– überflüssige Schritte eliminiert werden.

□

Satz 4.14

Sei $G = (V, \Sigma, R, S)$ eine KFG. Angenommen $A \overset{*}{\Rightarrow} x$ mit $x \in \Sigma^*$, dann liefert das rekursive Inferenzverfahren, dass $x \in \mathsf{L}(A)$.

Beweis. Wir zeigen den Satz mittels Induktion nach der Länge der Ableitung $A \overset{*}{\Rightarrow} x$.

1. Basis: Sei die Ableitung genau ein Schritt. Dann gilt $A \to x \in R$, also gilt $x \in \mathsf{L}(A)$ nach dem Basisfall des rekursiven Inferenzverfahrens.

2. Schritt: Angenommen $n + 1$ Schritte sind in der Ableitung notwendig:

$$A \Rightarrow X_1 X_2 \cdots X_n \overset{*}{\Rightarrow} x \; .$$

Wir können x als $x_1 x_2 \cdots x_n$ schreiben, wobei

– Wenn $X_i \in \Sigma$, dann $X_i = x_i$ oder

– $X_i \in V$, dann existiert eine Ableitung der Länge (maximal) n $X_i \overset{*}{\Rightarrow} x_i$. Nach Induktionshypothese folgt mit dem rekursiven Inferenzverfahren, dass $x_i \in \mathsf{L}(X_i)$.

Nach Annahme existiert eine Regel $A \to X_1 X_2 \cdots X_n \in R$. Somit folgt, dass das Wort $x_1 x_2 \cdots x_n$ in $\mathsf{L}(A)$ ist.

□

4.5. Anwendungen kontextfreier Grammatiken

Wir wenden uns nun kurz Anwendungen der Theorie der formalen Sprachen, bzw. präziser von regulären Sprachen und kontextfreien Grammatiken zu. Da die Anwendungsmöglichkeiten von Sprachen und Grammtiken vielfältig sind, konzentrieren wir uns dazu auf zwei Gebiete. In diesem Abschnitt skizzieren wir die klassische Anwendung von KFGs im *Compilerbau* [1], genauer in der Genenerierung eines Parsers mittels den C-Werkzeugen `lex` und `yacc`. Die Anwendung von KFGs in der *Wissensrepräsentation* wird im Anhang beschrieben, siehe Kapitel A.2.

Anwendungen von regulären Sprachen beziehungsweise regulären Ausdrücken[2] finden sich etwa auch in der Genese des Betriebssystems Unix. Reguläre Ausdrücke werden seit Beginn bei Unix verwendet. Beispiele hierfür sind `expr`, `awk`, `GNU Emacs`, `vi`, `lex` und `Perl`. Ken Thompson (1943–) baute reguläre Ausdrücke in den Texteditor `qed` ein und später in den Editor `ed`. Unter anderem für ihre Arbeiten zu Unix wurde Thompson und Dennis Ritchie (1941–2011) 1983 der *Turing Award*, der Nobelpreis der Informatik, verliehen.

Ein zentraler Teil bei Entwicklung eines Compilers sind *Parsergeneratoren*. Ein Parsergenerator verwandelt die Beschreibung einer formalen Sprache in einen Parser für diese Sprache. Parser

[2] Reguläre Ausdrücke stellen eine sehr intuitive Beschreibungsmöglichkeit von regulären Sprachen da, die aus Zeitgründen hier nicht behandelt werden können. Reguläre Ausdrücke werden jedoch in der Lehrveranstaltung Diskrete Strukturen behandelt, siehe auch [9, 11].

werden zur syntaktischen Analyse von Programmen verwendet. Zunächst wird mittels einer *lexikalische Analyse* der Eingabestrom aus `ASCII`-Zeichen analysiert und in terminale Symbole der formalen Sprache zerlegt, die man *Token* nennt.

Wir betrachten die Eingabe eines (simplen) Taschenrechners. Dann sind die Token die Ziffern der Rechung, charakterisiert etwa durch den folgenden regulären Ausdruck:

```
([0-9]+|([0-9]*\.[0-9]+))([eE][-+]?[0-9]+)?
```

Eine mögliche Eingabe, die diesem regulären Ausdruck entspriche wäre etwa: $1.0 + (2.1e1 - 3 * 5) * 0.5$. Dabei werden die Regeln der lexikalischen Analyse in der folgenden Form definiert:

Muster *Aktion*

Generell bestehen die Hilfsdateien des Parstergenerators aus (i) einem *Definitionsteil* (ii) einem *Regelteil*, der zwischen `%%` eingeschlossen wird, und (iii) einem *Programmteil* für benutzereigene Funktionen.

Beispiel 4.1: Arithmetische Ausdrücke

KFG G_2 für arithmetische Ausdrücke

$$E \to T \mid +T \mid -T \mid E + T \mid E - T$$
$$T \to F \mid T \cdot F \mid T/F$$
$$F \to \mathsf{c} \mid \mathsf{v} \mid (E)\ .$$

in G_2 gilt etwa $-c * v \in L(E)$:

	Wort x	Variable	Regel	Rekursion
1	c	F	$F \to \mathsf{c}$	
2	v	F	$F \to \mathsf{v}$	
3	c	T	$T \to F$	1
4	$\mathsf{c} \cdot \mathsf{v}$	T	$T \to T \cdot F$	3, 2
5	$-\mathsf{c} \cdot \mathsf{v}$	E	$E \to -T$	4

Beispiel 4.2: Regelteil calc.y (gekürzt)

```
expression:     term            { $$ = $1; }
        |       '+' term        { $$ = $2; }
        |       '-' term        { $$ = -$2; }
        |       expression '+' term { $$ = $1 + $3; }
        |       expression '-' term { $$ = $1 - $3; }
        ;

term:           factor                  { $$ = $1; }
        |       term '*' factor         { $$ = $1 * $3; }
        |       term '/' factor         { $$ = $1 / $3; }
        ;
```

```
factor:         NUMBER                  { $$ = $1; }
        |       '(' expression ')'      { $$ = $2; }
        ;
```

4.6. Zusammenfassung

In den 1940er und 1950er Jahren wurden von einigen Forschern einfachere Maschinen untersucht, die heute als „endliche Automaten“ bezeichnet werden. Diese Automaten, ursprünglich zur Simulation von Gehirnfunktionen von Warren McCulloch (1898–1969) und Walter Pitts (1923–1969) eingeführt, haben sich für verschiedene andere Zwecke als nützlich erwiesen. In den 1950er Jahren wurden die von McCulloch und Pitts vorgelegten Definition von Stephen Kleene (1909–1994) aufgenommen und mathematisch präzise gefasst. Auf Kleene geht die Definition von *regulären Sprachen* zurück und in seinem Namen wird der Operator * auch oft als *Kleene-Stern* bezeichnet. In den späten 1950er Jahren begann zudem der Linguist Noam Chomsky (1928–), *formale Grammatiken* zu untersuchen. Diese Grammatiken dienen heute als Grundlage einiger wichtiger Softwarekomponenten, wie etwa Compilern.

In der technischen Informatik kommen neben den hier betrachteten Formen endlicher Automaten auch endliche Automaten nach *Mealy* oder *Moore* zur Anwendung. Diese Automaten wurden von George Mealy (1927–2010) beziehungsweise Edward Moore (1925—2003) eingeführt. Mealy und Moore Automaten verfügen über ein separates Ausgabealphabet und können direkt zur Beschreibung von Funktionen verwandt werden. Dies führt jedoch nicht zu einer Steigerung der prinzipiellen Ausdrucksstärke.

Die Konstruktion von Compilern wurde vor allem von Alfred Aho (1941–) und Jeffrey Ullman (1942–) auf eine solide formale Grundlage gestellt, siehe [1]. Diese grundlegende und praktisch äußerst relevanten Arbeit wurde kürzlich mit der Verleihung des Turingaward für die beiden belohnt.[3]

4.7. Aufgaben

Übungsaufgabe 4.1

Betrachten Sie das Alphabet $\Sigma = \{a, b\}$ sowie die formalen Sprachen $L = \{aa, b\}$ und $M = \{\epsilon, a, bb\}$ über Σ.

1. Berechnen Sie Σ^0.
2. Berechnen Sie Σ^3.
3. Berechnen Sie $L \cup M^2$
4. Berechnen Sie LM^2
5. Berechnen Sie $(LM)^2$
6. Beschreiben Sie L^+ in Worten.

[3] Siehe https://de.wikipedia.org/wiki/Turing_Award.

7. Beschreiben Sie L^* in Worten.

Übungsaufgabe 4.2

Betrachten Sie die rekursive Definition 4.3 der Konkatenation und die folgende alternative Definition (@) von Konkatenation, wobei x und y Wörter, sodass $x = a_1a_2\cdots a_m$, $y = b_1b_2\cdots b_n$:

$$x@y = a_1a_2\cdots a_mb_1b_2\cdots b_n\ .$$

Zeigen Sie, dass beide Definitionen äquivalent sind.

Übungsaufgabe 4.3

Betrachten Sie die Grammatik $G = (V, \Sigma, R, S)$ mit $V = \{S, X, Y\}$, $\Sigma = \{a, b, c\}$, und R gegeben durch die Regeln

$$\begin{aligned} S &\to \epsilon \mid SS \mid X \\ X &\to YY \\ Y &\to aY \mid b \\ YaY &\to c \end{aligned}$$

Welche der folgenden Wörter können von S abgeleitet werden? Geben Sie eine Ableitung an, wenn möglich.

1. ϵ
2. a
3. bb
4. c

Lösung.

1. $S \Rightarrow \epsilon$.
2. Es kann keine Ableitung $S \overset{*}{\Rightarrow} a$ geben. Die einzige Möglichkeit ein a zu bekommen ist mittels der Regel $Y \to aY$. Das Y in der Konklusion liefert schlussendlich (zumindest) ein b oder ein c, somit gilt $a \notin \mathsf{L}(G)$.
3. $S \Rightarrow X \Rightarrow YY \Rightarrow Yb \Rightarrow bb$.
4. $S \Rightarrow X \Rightarrow YY \Rightarrow YaY \Rightarrow c$.

Übungsaufgabe 4.4

Sei

$$L = \{w \in \Sigma^* \mid w \text{ hat die selbe Anzahl von 0en und 1en}\}$$

eine formale Sprache über $\Sigma = \{0, 1\}$. Finden Sie eine Grammatik G, welche L erzeugt.

Lösung. Wir wählen $G = (\{S\}, \Sigma, R, S)$ mit den Regeln R

$$S \to \epsilon \mid S0S1S \mid S1S0S$$

Wenn $S \overset{*}{\Rightarrow} x$, dann $x \in L$ ist offensichtlich. Der Beweis für "Wenn $x \in L$, dann $S \overset{*}{\Rightarrow} x$" erfolgt mittels Induktion über $|x|$. Im Basisfall ist $|x| = 0$, somit $x = \epsilon$. Da $S \overset{*}{\Rightarrow} \epsilon$ gilt der Basisfall. Im Schrittfall ist $|x| > 0$. Die Induktionshypothese besagt, dass es für alle Wörter $w \in L$ mit $|w| < |x|$ eine Ableitung $S \overset{*}{\Rightarrow} w$ gibt. Wir unterscheiden vier Fälle (beachte $|x| \geq 2$):

1. $x = 0w0$. Da $x \in L$ muss es $x_1, x_2, x_3 \in L$ geben mit $x = 0x_1 1x_2 1x_3 0$. Wir erhalten (wobei für $1 \leq i \leq 3$ die IH $S \overset{*}{\Rightarrow} x_i$ liefert):

$$\begin{aligned} S \Rightarrow S0S1S \Rightarrow 0S1S \overset{*}{\Rightarrow} 0x_1 1S \Rightarrow 0x_1 1S1S0S \overset{*}{\Rightarrow} 0x_1 1x_2 1S0S \\ \overset{*}{\Rightarrow} 0x_1 1x_2 1x_3 0S \Rightarrow 0x_1 1x_2 1x_3 0 = x \end{aligned}$$

2. $x = 0w1$. Da $x \in L$ muss auch $w \in L$. Aus der IH folgt $S \overset{*}{\Rightarrow} w$ und somit $S \Rightarrow S0S1S \Rightarrow 0S1S \Rightarrow 0S1 \overset{*}{\Rightarrow} 0w1 = x$.

3. $x = 1w0$. Analog zu (ii).

4. $x = 1w1$. Analog zu (i).

Übungsaufgabe 4.5

Sei $G = (V, \Sigma, R, S)$ eine Grammatik mit der Eigenschaft, dass für alle Regeln $P \to Q$ gilt:

(1) $P \in V$

(2) $Q = \epsilon$ oder $Q = a$ oder $Q = aX$ (wobei $a \in \Sigma$, $X \in V$).

1. Begründen Sie, dass jede Grammatik der obigen Gestalt rechtslinear ist.
2. Kann jede von einer rechtslinearen Grammatik erzeugte Sprache auch von einer Grammatik mit der obigen Gestalt erzeugt werden? (Wenn ja: wie? Wenn nein: warum nicht?)

Übungsaufgabe 4.6

Betrachten Sie die Grammatik $G = (\{S, X\}, \Sigma, R, S)$ mit Regeln R

$$\begin{aligned} S &\to X0 \mid X1 \mid 00S \mid 11S \\ X &\to X0 \mid X1 \mid 0 \mid 1 \end{aligned}$$

1. Beschreiben Sie $\mathsf{L}(G)$ in Worten.

2. Welche Eigenschaften (laut Definition 3.11) hat Ihre Grammatik?
3. Welchen Typ hat $\mathsf{L}(G)$?

Lösung.

1. Die letzten beiden Regeln sind redundant. $\mathsf{L}(G)$ beinhaltet alle Wörter mit Länge mindestens zwei bestehend aus Nullen und Einsen.
2. G ist kontextfrei.
3. Typ 3 (obwohl G kontextfrei.)

Übungsaufgabe 4.7

Sei $\Sigma = \{0, 1\}$. Geben Sie einen DEA $A = (Q, \Sigma, \delta, s, F)$ an, sodass

$$\mathsf{L}(A) = \{x \in \Sigma^* \mid \text{die Anzahl der Nullen in } x \text{ ist ein Vielfaches von } 3\}$$

Beispiele: $\epsilon \in \mathsf{L}(A)$, $0 \notin \mathsf{L}(A)$, $1 \in \mathsf{L}(A)$, $11 \in \mathsf{L}(A)$, $101 \notin \mathsf{L}(A)$, $0100 \in \mathsf{L}(A)$.
Hinweis: Geben Sie den DEA über die Zustandstabelle sowie den Zustandsgraphen an.

Übungsaufgabe 4.8

Sei A ein DEA. Welche der folgenden Aussagen ist wahr (allgemein gültig)?

1. A hat genau einen Startzustand.
2. A hat genau einen akzeptierenden Zustand.
3. Das Eingabealphabet von A hat genau einen Buchstaben.
4. Der Startzustand von A kann kein akzeptierender Zustand sein.
5. A akzeptiert eine Sprache, die kontextfrei, aber nicht regulär ist.
6. Alle anderen Aussagen sind falsch.

Lösung. A hat genau einen Startzustand.

Übungsaufgabe 4.9

Sei $\Sigma = \{0, \ldots, 9\}$.

1. Geben Sie eine kontextfreie Grammatik G an, welche die natürlichen Zahlen $\mathbb{N} = \{0, 1, 2, \ldots\}$ erzeugt.
   Hinweis: Achten Sie darauf, dass Ihre Grammatik keine Wörter mit führenden Nullen erzeugt, d.h. $12 \in \mathsf{L}(G)$, aber $0012 \notin \mathsf{L}(G)$.
2. Geben Sie eine Linksableitung und eine Rechtsableitung für das Wort 123 an.

3. Ist Ihre Grammatik eindeutig?

Lösung. Wir geben nur die Grammatik an. Setze $G = (\{S, N, D, P\}, \{0, \dots, 9\}, R, S)$ mit Regeln

$$\begin{aligned} S &\to 0 \mid PN \\ N &\to \epsilon \mid DN \\ D &\to 0 \mid P \\ P &\to 1 \mid \cdots \mid 9 \end{aligned}$$

Übungsaufgabe 4.10

Betrachten Sie die induktive Definition von Palindromen sowie die Grammatik G_1 aus der Vorlesung. Beweisen Sie: Wenn x ein Palindrom, dann $x \in \mathsf{L}(G_1)$ (d.h. $P \overset{*}{\underset{G_1}{\Rightarrow}} x$).

Übungsaufgabe 4.11

Betrachten Sie die (kontextfreie) Grammatik $G_2 = (\{S\}, \{(,)\}, R, S)$ mit Regeln R

$$S \to \epsilon \mid (S) \mid SS$$

und das Wort $w = (()(())) \in \mathsf{L}(G_2)$.

1. Geben Sie eine Linksableitung für w an.
2. Geben Sie eine Rechtsableitung für w an.
3. Sind alle Ableitungen für w entweder eine Links- oder Rechtsableitung?
4. Leiten Sie $S \overset{*}{\underset{G_2}{\Rightarrow}} w$ mittels rekursiver Inferenz ab.
5. Geben Sie einen Syntaxbaum für G_2 mit Wurzel P und Ergebnis w an.
6. Ist G_2 eindeutig?

Lösung.

3. Nein, bei $S \Rightarrow (S) \Rightarrow (SS) \Rightarrow (S(S)) \Rightarrow (()(S)) \Rightarrow (()(()))$ wechselt man zwischen links und rechts.
6. Nein, da $S \overset{*}{\underset{G_2}{\Rightarrow}} \epsilon$ und $S \overset{*}{\underset{G_2}{\Rightarrow}} SS \overset{*}{\underset{G_2}{\Rightarrow}} S \overset{*}{\underset{G_2}{\Rightarrow}} ()$.

Übungsaufgabe 4.12

Beweisen oder widerlegen Sie:

> Es gibt eine formale Sprache L vom Typ 3, für die es keine kontextfreie Grammatik G gibt, welche L erzeugt.

Lösung. Die Behauptung ist falsch, da jede rechtslineare Grammatik kontextfrei ist. Rechtslineare Grammatiken erzeugen genau die Klasse der Sprachen vom Typ 3.

Teil III.

Berechenbarkeit, Komplexität & Verifikation

5.

Einführung in die Berechenbarkeitstheorie

In diesem Kapitel wird das formale Modell des endlichen Automaten zu *Turing-vollständigen Berechenbarkeitsmodellen* erweitert und eine Einführung in die *Berechenbarkeitstheorie* gegeben. In Abschnitt 5.1 liegt unser Hauptaugenmerk auf der Klärung der Frage welche Probleme prinzipiell algorithmisch lösbar sind. Das Berechnungsmodell der *Turingmaschinen* wird in Abschnitt 5.2 eingeführt. Diesem Modell werden in Abschnitt 5.3 Registermaschinen gegenübergestellt.

Schließlich gehen wir in Abschnitt 5.5 kurz auf die Bedeutung von Berechnungsmodellen für die Informatik ein und stellen die betrachteten Konzepte in einen historischen Kontext. Außerdem finden sich in Abschnitt 5.6 (optionale) Aufgaben zu den Themenbereichen dieses Kapitels, die zur weiteren Vertiefung dienen sollen.

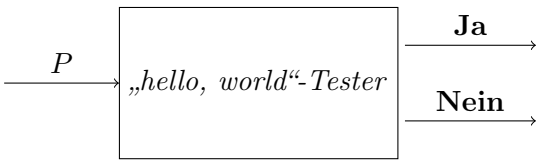


Abbildung 5.1.: Ein hypothetischer „hello, world“-Tester

5.1. Algorithmisch unlösbare Probleme

In diesem Abschnitt wollen wir uns einführend mit der Frage beschäftigen, ob alle Probleme algorithmisch lösbar sind. Hier bedeutet „algorithmisch lösbar“, dass es einen *Algorithmus*, das heißt ein Programm gibt, welches das Problem vollständig, das heißt auf allen möglichen Eingaben, löst. Leider müssen wir diese Frage mit „Nein“ beantworten. In der Folge erklären wir, warum diese Antwort nicht wirklich überraschend ist. Zunächst betrachten wir ein sehr einfaches C-Programm P:

```
int main(void) {
    printf("hello, world");
}
```

Wie leicht einzusehen, gibt P die Worte „hello, world“ aus und terminiert. In der Folge nennen wir jedes Programm, das die Zeichenreihe „hello, world“ als die ersten 12 Buchstaben seiner Ausgabe druckt ein *„hello, world“-Programm*. Wir setzen dabei nicht voraus, dass das Programm tatsächlich auf seiner Eingabe hält.

Nun untersuchen wir, ob es möglich ist, ein Programm zu schreiben, das testet, ob ein bestimmtes Programm ein „hello, world“-Programm ist. Schematisch können wir einen hypothetisch angenommenen *„hello, world“-Tester* H wie in Abbildung 5.1 beschreiben. Der Tester H erhält als Eingabe ein Programm P und antwortet entweder mit „Ja“, wenn P ein „hello, world“-Programm ist und sonst mit „Nein“. In Anbetracht der Einfachheit des Programms P erscheint diese Aufgabe recht einfach. Diese Einfachheit ist jedoch trügerisch. Dazu betrachten wir die folgende Variante P_1 des Programms P. Wir setzen die Funktion `exp` voraus, wobei $\texttt{exp}(x, y) = x^y$.

```
int main(void) {
    int n, sum, x, y, z;
    scanf("%d", &n);
    sum = 3;
    while (1) {
        for (x = 1; x <= sum - 2; x++)
            for (y = 1; y <= sum - x - 1; y++) {
                z = sum - x - y;
                if (exp(x, n) + exp(y, n) == exp(z, n))
                    printf("hello, world");
            }
        sum++;
    }
}
```

Es ist nicht schwer einzusehen, dass P_1 die Eingabe einer natürlichen Zahl n erwartet und dann prüft, ob es natürliche Zahlen x, y und z gibt ($x, y, z \geqslant 1$), sodass die Gleichung

$$x^n + y^n = z^n \,, \tag{5.1}$$

wahr wird. Wenn eine solche Lösung gefunden wird, wird der String „hello, world“ gedruckt. Ist das Programm P_1 ein „hello, world“-Programm? Angenommen als Eingabe setzen wir $n = 2$. Dann ist leicht zu überprüfen, dass $x = 3$, $y = 4$ und $z = 5$ die Gleichung (5.1) löst. Das heißt für $n = 2$ ist P_1 ein „hello, world“-Programm. Genauer sehen wir, dass P_1 „hello, world“ druckt, sobald das erste *pythagoreische Tripel* gefunden wird. Was passiert nun für $n \geqslant 3$? Dann müssen wir feststellen,

dass das Programm niemals den String „hello, world“ schreibt, da die Gleichung (5.1) für $n \geqslant 3$ unlösbar ist. Dieser Sachverhalt wurde von Pierre de Fermat (1607–1665) im 17. Jahrhundert vermutet. Es dauerte über 300 Jahre bis der britische Mathematiker Andrew Wiles (1953–) im Jahr 1995 diese Vermutung auch tatsächlich beweisen konnte [14]. Es hat also der Anstrengung von Generationen von Mathematikern und Mathematikerinnen bedurft, um festzustellen, dass das Programm P_1 in bestimmten Fällen *kein* „hello, world“-Programm ist.

Wir betrachten eine weitere Variante P_2 von P. In P_2 setzten wir die Boolesche Funktion `primes` voraus, wobei $\texttt{primes}(n) = 1$ gdw. n eine Primzahl ist.[1]

```
int main(void) {
    int sum = 4, x, y, test;
    while (1) {
        test = 1;
        for (x = 2; x <= sum; x++) {
            y = sum - x;
            if (primes(x) && primes(y))
                test = 0;
        }
        if (test)
            printf("hello, world");
        sum = sum + 2;
    }
}
```

Es ist nicht schwer einzusehen, dass das Programm P_2 ein „hello, world“-Programm ist, wenn eine gerade natürliche Zahl größer als 2 existiert, die nicht als Summe zweier Primzahlen geschrieben werden kann. Ob es überhaupt möglich ist, jede gerade natürliche Zahl größer als 2 als Summe zweier Primzahlen zu schreiben, ist zur Zeit nicht bekannt. Christian Goldbach (1690–1764) hat diese Vermutung, die deshalb *Goldbachsche Vermutung* genannt wird, im 18. Jahrhundert aufgestellt.

Diese beiden Varianten des anfänglich betrachteten „hello, world“-Programms zeigen uns, dass die Konstruktion eines „hello, world“-Testers keineswegs eine einfache Angelegenheit ist. In der Tat kann man beweisen, dass es keinen „hello, world“-Tester H geben kann. Wir formulieren allgemein das folgende Problem:

Problem

Gegeben ein beliebiges Programm P. Ist P ein „hello, world“-Programm?

Dieses Problem ist *algorithmisch nicht lösbar*, da wir keinen Algorithmus dafür angeben können. Probleme, die in diesem Sinn nicht gelöst werden können, nennen wir *unentscheidbar*. Wie kann die algorithmische Unlösbarkeit formal gezeigt werden, wie kommt man also zu der Behauptung, dass etwas nicht algorithmisch lösbar ist? Dazu bräuchte man im Prinzip eine formale Definition was ein „Algorithmus“ ist. Das ist jedoch nicht möglich, da ein „Algorithmus“ ein intuitives Konzept ist. Allerdings hat man sogenannte *abstrakte Berechnungsmodelle* studiert, die in einer geeigneten Weise alle möglichen Beschreibungen von Algorithmen darstellen können, die man sich bis jetzt vorstellen konnte.

[1] Eine solche Funktion wird auch *Primzahltest* genannt. Das Testen, ob eine bestimmte natürliche Zahl eine Primzahl ist, ist entscheidbar.

Im weiteren Verlauf dieses Kapitels werden wir zwei solche Modelle studieren: *Turingmaschinen* und *Registermaschinen*. Es kann gezeigt werden, dass diese Modelle äquivalent sind. Jedes Programm, dass auf einer Turingmaschine läuft, kann in ein Programm einer Registermaschine umgeschrieben werden und umgekehrt. Ähnliches gilt für alternative Berechnungsmodelle wie etwa Grammatiken, den von Alonzo Church eingeführten *λ-Kalkül* [3] oder *Termersetzungssysteme* [2]: Alle untersuchten Präzisierungen des Begriffs „Algorithmus“ beschreiben exakt die gleiche Menge von Programmen. Diese Beobachtung hat schon in den 1930er Jahren die Grundlage für die so genannte *Church-Turing-These* geliefert:

These

Jedes algorithmisch lösbare Problem ist auch mit Hilfe einer Turingmaschine lösbar.

Wir schließen diesen Abschnitt mit einer (sehr) unvollständigen Liste unentscheidbarer Probleme. Die folgenden Probleme sind *unentscheidbar*:

- Das Problem, ob ein beliebiges Programm auf seiner Eingabe hält. (*Halteproblem*)
- Postsches Korrespondenzproblem (*PCP*): Gegeben zwei Listen von Wörtern

  $$x_1, \ldots, x_n \quad \text{und} \quad y_1, \ldots, y_n \; .$$

  Gesucht sind Indizes $i_1, i_2, \ldots, i_m$ (nicht notwendigerweiße verschieden), sodass

  $$x_{i_1} \ldots x_{i_m} = y_{i_1} \ldots y_{i_m} \; .$$

- Das Problem, ob eine beliebige kontextfreie Grammatik eindeutig ist.

5.2. Turingmaschinen

Im Vergleich zu anderen Konzepten zur Beschreibung der Klasse der berechenbaren Funktionen, stellen Turingmaschinen eines der einfachsten abstrakten Berechnungsmodelle dar. Wir beschreiben hier nur deterministische, 1-Band-Turingmaschinen. Äquivalente Formulierungen von Turingmaschinen, wie etwa Maschinen mit mehreren Bändern, mehreren Leseköpfen oder nichtdeterministische Turingmaschinen werden in der Vorlesung Diskrete Strukturen behandelt.

Eine *Turingmaschine* (abgekürzt *TM*) besteht aus einer endlichen Anzahl von Zuständen Q, einem einseitig unendlichen Band und einem Lese- und Schreibkopf, der eine Position nach links oder rechts wechseln kann und Symbole lesen beziehungsweise schreiben kann. Das einseitig unendliche Band ist auf der linken Seite durch $\vdash$ begrenzt und unbeschränkt auf der rechten Seite. Abbildung 5.2 liefert einen schematisierten Überblick.

Definition 5.1: Turingmaschine

Eine *deterministische, einbändige Turingmaschine* M ist ein 9-Tupel

$$M = (Q, \Sigma, \Gamma, \vdash, \sqcup, \delta, s, t, r) \; ,$$

sodass

1. Q eine endliche Menge von *Zuständen*,

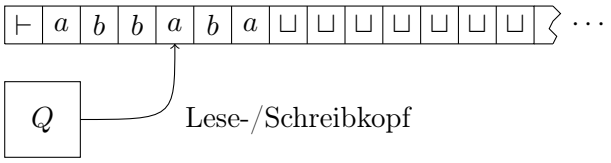


Abbildung 5.2.: Schema einer Turingmaschine

2. $\Sigma \subseteq \Gamma$ eine endliche Menge von *Eingabesymbolen*,
3. Γ eine endliche Menge von *Bandsymbolen*,
4. $\sqcup \in \Gamma \setminus \Sigma$, das *Blanksymbol*,
5. $\vdash \in \Gamma \setminus \Sigma$, der *linke Endmarker*,
6. $\delta\colon Q \times \Gamma \to Q \times \Gamma \times \{\mathrm{L}, \mathrm{R}\}$ die *Übergangsfunktion*,
7. $s \in Q$, der *Startzustand*,
8. $t \in Q$, der *akzeptierende Zustand* und
9. $r \in Q$, der *verwerfende Zustand* mit $t \neq r$.

Informell bedeutet $\delta(p, a) = (q, b, d)$: „Wenn die TM M im Zustand p das Symbol a liest, dann ersetzt M das Zeichen a durch das Zeichen b, der Lese-/Schreibkopf bewegt sich einen Schritt in die Richtung d und M wechselt in den Zustand q.“ Wir verlangen, dass das Symbol $\vdash$ niemals überschrieben werden kann und die Maschine niemals über die linke Begrenzung hinaus fährt. Dies wird formal durch die folgende Bedingung festgelegt: Für alle $p \in Q$, existiert $q \in Q$ mit:

$$\delta(p, \vdash) = (q, \vdash, \mathrm{R})\ . \tag{5.2}$$

Außerdem verlangen wir dass die Maschine, sollte sie den akzeptierenden beziehungsweise verwerfenden Zustand erreicht haben, diesen nicht mehr verlassen kann. Das heißt für alle $b \in \Gamma$ existieren $c, c' \in \Gamma$ und $d, d' \in \{\mathrm{L}, \mathrm{R}\}$ sodass gilt:

$$\delta(t, b) = (t, c, d) \tag{5.3}$$
$$\delta(r, b) = (r, c', d') \tag{5.4}$$

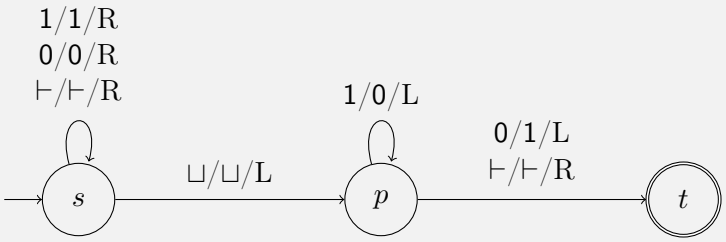
Die Zustandsmenge Q und die Übergangsfunktion δ einer TM M wird auch als die *endliche Kontrolle* von M bezeichnet.

Beispiel 5.1

Sei $M = (\{s, p, t, r\}, \{0, 1\}, \{\vdash, \sqcup, 0, 1\}, \vdash, \sqcup, \delta, s, t, r)$ eine Turingmaschine wobei die Übergangsfunktion δ durch die *Zustandstabelle*

	$\vdash$	0	1	$\sqcup$
s	$(s, \vdash, \mathrm{R})$	$(s, 0, \mathrm{R})$	$(s, 1, \mathrm{R})$	$(p, \sqcup, \mathrm{L})$
p	$(t, \vdash, \mathrm{R})$	$(t, 1, \mathrm{L})$	$(p, 0, \mathrm{L})$	·

oder durch das *Zustandsübergangsdiagramm*



angegeben werden kann.
In der Tabelle sowie im Zustandsübergangsdiagramm wurde auf die Übergänge für die Zustände t und r verzichtet. Diese können so gewählt werden, dass der Zustand nicht geändert wird, das gelesene Zeichen wieder geschrieben wird und sich der Schreib-/Lesekopf nach rechts bewegt. Trivialerweise sind die Bedingungen (5.3)–(5.4) dann erfüllt. Ebenso kann der fehlende Übergang für $\delta(p, \sqcup)$ gewählt werden.

Beachten Sie, dass eine Turingmaschine M nach Definition 5.1 niemals zur Ruhe kommt. Zwar kann M in ihren akzeptierenden oder verwerfenden Zustand wechseln und diesen dann auch nicht mehr verlassen, aber M bleibt trotzdem immer in Bewegung. Dennoch sprechen wir vom *Halten* der TM M, wenn M entweder den Zustand t oder r erreicht, andernfalls sagen wir: M *hält nicht* (oder auch *terminiert* nicht).

Zu jedem Zeitpunkt enthält das Band einer TM M ein unendliches Wort der Form $y\sqcup^\infty$, wobei $\sqcup^\infty$ das unendliche Wort

$$\sqcup\sqcup\sqcup\sqcup\sqcup\sqcup\sqcup\sqcup\sqcup\sqcup\cdots ,$$

bezeichnet. Obwohl das Wort $y\sqcup^\infty$ unendlich ist, ist es endlich repräsentierbar, da nur die Darstellung von y von Interesse ist und $|y| \in \mathbb{N}$.

Definition 5.2: Konfiguration einer TM

Eine *Konfiguration* einer TM M ist ein Tripel (p, z, n), sodass

- $p \in Q$ der aktuelle Zustand,
- $z = y\sqcup^\infty$ der aktuelle Bandinhalt ($y \in \Gamma^*$) und
- $n \in \mathbb{N}$ die Position des Lese-/Schreibkopfes am Band.

Die *Startkonfiguration* bei Eingabe $x \in \Sigma^*$ ist die Konfiguration

$$(s, \vdash x\sqcup^\infty, 0) .$$

In der Folge definieren wir eine binäre Relation zwischen Konfigurationen, um einen Rechenschritt einer Turingmaschine konzise formalisieren zu können.

Definition 5.3

Sei $z \in \Gamma^*$. Wir schreiben z_n für das n-te Symbol des Wortes z. Die Relation $\xrightarrow[M]{1}$ ist wie folgt definiert:

$$(p, z, n) \xrightarrow[M]{1} \begin{cases} (q, z', n-1) & \text{wenn } \delta(p, z_n) = (q, b, \mathrm{L}) \\ (q, z', n+1) & \text{wenn } \delta(p, z_n) = (q, b, \mathrm{R}) \end{cases}$$

Hier bezeichnet z' das Wort, das wir aus z erhalten, wenn z_n durch b ersetzt wird.

Wir verwenden griechische Buchstaben vom Anfang des Alphabets um Konfigurationen zu bezeichnen.

Definition 5.4

Wir definieren die reflexive, transitive Hülle $\xrightarrow[M]{*}$ von $\xrightarrow[M]{1}$ induktiv:

1. $\alpha \xrightarrow[M]{0} \alpha$
2. $\alpha \xrightarrow[M]{i+1} \beta$, wenn $\alpha \xrightarrow[M]{i} \gamma \xrightarrow[M]{1} \beta$ für eine Konfiguration γ und
3. $\alpha \xrightarrow[M]{*} \beta$, wenn $\alpha \xrightarrow[M]{i} \beta$ für ein $i \geqslant 0$.

Beispiel 5.2

Für die TM M aus Beispiel 5.1 gilt

$$(s, \vdash 0010 \sqcup^\infty, 0) \xrightarrow[M]{*} (t, \vdash 0011 \sqcup^\infty, 3)\,.$$

Lösung.

$$(s, \vdash 0010 \sqcup^\infty, 0) \xrightarrow[M]{*} (s, \vdash 0010 \sqcup^\infty, 5) \xrightarrow[M]{1} (p, \vdash 0010 \sqcup^\infty, 4) \xrightarrow[M]{1} (t, \vdash 0011 \sqcup^\infty, 3)$$

Definition 5.5: Sprache einer TM

Sei $M = (Q, \Sigma, \Gamma, \vdash, \sqcup, \delta, s, t, r)$ eine TM. Die Turingmaschine M *akzeptiert* die Eingabe $x \in \Sigma^*$, wenn gilt

$$(s, \vdash x \sqcup^\infty, 0) \xrightarrow[M]{*} (t, y, n)\,,$$

für ein $y \in \Gamma^*$ und $n \in \mathbb{N}$. M *verwirft* x, wenn

$$(s, \vdash x \sqcup^\infty, 0) \xrightarrow[M]{*} (r, y, n)\,,$$

für ein $y \in \Gamma^*$ und $n \in \mathbb{N}$. Wir sagen M *hält* bei Eingabe x, wenn M die Eingabe x entweder akzeptiert oder verwirft. Andernfalls *hält* M auf x *nicht*. Eine TM M heißt *total*, wenn sie auf allen Eingaben hält. Die Menge $\mathsf{L}(M)$ bezeichnet die Menge aller von M akzeptierten Wörter.

Beispiel 5.3

Für die TM

$$M = (\{s, q_0, q_1, q_0', q_1', q, t, r\}, \{0, 1\}, \{\vdash, \sqcup, 0, 1\}, \vdash, \sqcup, \delta, s, t, r)$$

mit δ gegeben durch die Zustandstabelle

	$\vdash$	0	1	$\sqcup$
s	$(s, \vdash, R)$	$(q_0, \vdash, R)$	$(q_1, \vdash, R)$	$(t, \sqcup, L)$
q_0	·	$(q_0, 0, R)$	$(q_0, 1, R)$	$(q_0', \sqcup, L)$
q_1	·	$(q_1, 0, R)$	$(q_1, 1, R)$	$(q_1', \sqcup, L)$
q_0'	$(r, \vdash, R)$	$(q, \sqcup, L)$	$(r, \sqcup, L)$	·
q_1'	$(r, \vdash, R)$	$(r, \sqcup, L)$	$(q, \sqcup, L)$	·
q	$(s, \vdash, R)$	$(q, 0, L)$	$(q, 1, L)$	·

ist $\mathsf{L}(M) = \{w \in \{0, 1\}^* \mid w \text{ ist ein Palindrom gerader Länge}\}$.

Der folgende Satz zeigt, dass Turingmaschinen die gleichen Sprachen beschreiben können wie Grammatiken. Für den Beweis des Satzes sei auf [6] verwiesen.

Satz 5.1

Sei M eine Turingmaschine. Dann ist $\mathsf{L}(M)$ rekursiv aufzählbar (vergleiche Definition 4.12). Umgekehrt gibt es zu jeder rekursiv aufzählbaren Sprache L eine Turingmaschine M mit $L = \mathsf{L}(M)$. □

Eine Sprache L heißt *rekursiv*, wenn es eine *totale* TM M gibt mit $L = \mathsf{L}(M)$. Die Klasse der rekursiven Sprachen ist echt größer als die Klasse der beschränkten Sprachen, aber echt kleiner als die Klasse der rekursiv aufzählbaren Sprachen [6].

In der Programmierung nennt man Algorithmen rekursiv, die sich selbst aufrufen. Hier wird die selbe Bezeichnung für ein anderes Konzept verwendet.[2]

Um Turingmaschinen mit Registermaschinen, die wir im nächsten Abschnitt einführen werden, vergleichen zu können, definieren wir neben der Sprache, die von einer TM akzeptiert wird, wie eine partielle Funktion $f\colon \mathbb{N}^k \to \mathbb{N}$ von einer TM berechnet werden kann. Dazu repräsentieren wir natürliche Zahlen *unär*: Eine Zahl $n \in \mathbb{N}$ wird auf dem Band der TM durch n Wiederholungen des Zeichens $\sqcap$ dargestellt. Wir schreiben abkürzend $\sqcap^n$ für n-maliges Hinschreiben von $\sqcap$.

Definition 5.6: Berechenbarkeit mit einer TM

Sei

$$M = (Q, \{\sqcap, \square\}, \{\vdash, \sqcup, \sqcap, \square\}, \vdash, \sqcup, \delta, s, t, r)\ ,$$

eine TM. Eine partielle Funktion $f\colon \mathbb{N}^k \to \mathbb{N}$ heißt *M-berechenbar*, wenn gilt

$$f(n_1, \ldots, n_k) = m \quad \text{gdw.} \quad (s, \vdash\ \sqcap^{n_1}\square \cdots \square\, \sqcap^{n_k}\sqcup^\infty, 0) \xrightarrow[M]{*} (t, \vdash\ \sqcap^m\sqcup^\infty, n)\ .$$

[2] Dennoch gibt es einen Zusammenhang. Rekursive Funktionen können nämlich ebenfalls genau die Klasse von rekursiven Mengen, d.h., entscheidbare Probleme repräsentieren.

Hier dient das Zeichen □ zur Trennung der unären Repräsentation. Eine partielle Funktion $f\colon \mathbb{N}^k \to \mathbb{N}$ heißt *berechenbar mit einer TM*, wenn eine TM M über dem Alphabet $\{\sqcap, \square\}$ existiert, sodass f M-berechenbar ist.

Beachten Sie, dass die in Definition 5.6 definierte partielle Funktion undefiniert ist, wenn die TM M nicht hält.

5.3. Registermaschinen

Eine *Registermaschine* (abgekürzt *RM*) ist eine Maschine, die eine endliche Anzahl von Registern, $x_1, \ldots, x_n$ besitzt. Die Register enthalten beliebig große natürliche Zahlen. Eine RM führt ein Programm P aus, dessen Befehle an eine stark vereinfachte imperative Sprache erinnern. Es gibt verschiedene, äquivalente Möglichkeiten, die Instruktionen einer RM zu wählen (siehe etwa [6]). Wir werden als Programme so genannte while-Programme verwenden. Diese Programme verwenden im Befehlssatz neben einfachen Zuweisungen auch bedingte Schleifenaufrufe, also while-*Schleifen*. Eine andere Möglichkeit wären etwa goto-Programme. In goto-Programmen werden die Befehle durchnummeriert und es kommen zu den einfachen Zuweisungen bedingte Sprunganweisungen, goto-*Befehle* hinzu [6, 13]. Da wir uns auf while-Programme beschränken, sprechen wir der Einfachheit halber schlicht von Programmen.

Definition 5.7: Registermaschine

Eine *Registermaschine* R ist ein Paar

$$R = ((x_i)_{1 \leqslant i \leqslant n}, P) \ ,$$

sodass $(x_i)_{1 \leqslant i \leqslant n}$ eine Sequenz von n *Registern* x_i ist und P ein Programm. *Programme* sind endliche Folgen von Befehlen und sind induktiv definiert:

1. Für jedes Register x_i sind die folgenden Instruktionen sowohl Befehle wie Programme:

$$x_i := x_i + 1 \qquad x_i := x_i - 1 \ .$$

2. Wenn P_1, P_2 Programme sind, dann ist

$$P_1; P_2 \ ,$$

ein Programm und

$$\textsf{while } x_i \neq 0 \textsf{ do } P_1 \textsf{ end} \ ,$$

ist sowohl ein Befehl als auch ein Programm.

Wir beschreiben die Semantik eines Programms nur informell. Für eine Registermaschine $R = ((x_i)_{1 \leqslant i \leqslant n}, P)$ bedeuten die Befehle

$$x_i := x_i + 1 \qquad x_i := x_i - 1 \ ,$$

dass der Inhalt des Register x_i entweder um 1 erhöht oder vermindert wird, wobei im zweiten Fall

$x_i > 0$ gelten muss. Der Befehl $x_i := x_i - 1$ angewandt auf ein Register $x_i = 0$, verändert den Registerinhalt nicht. Das Programm $P_1; P_2$ bedeutet, dass zunächst das Programm P_1 und dann das Programm P_2 ausgeführt wird. Schließlich bedeutet der Befehl (und das Programm)

$$\mathsf{while}\ x_i \neq 0\ \mathsf{do}\ P_1\ \mathsf{end}\ ,$$

dass der Schleifenrumpf P_1 solange ausgeführt werden soll bis die Bedingung $x_i \neq 0$ falsch wird. Das Ende eines Programms ist erreicht, wenn kein nächster auszuführender Befehl existiert. In diesem Fall *hält* die Registermaschine auf ihrer Eingabe.

Definition 5.8: Berechenbarkeit mit einer RM

Sei $R = ((x_i)_{1 \leqslant i \leqslant n}, P)$ eine Registermaschine. Eine partielle Funktion $f\colon \mathbb{N}^k \to \mathbb{N}$, heißt R-berechenbar, wenn gilt

$$f(n_1, \ldots, n_k) = m \quad \text{gdw.} \quad \begin{array}{l} R \text{ mit } n_i \text{ in den Registern } x_i \text{ für } 1 \leqslant i \leqslant k \text{ startet und} \\ \text{die Programmausführung mit } m \text{ im Register } x_{k+1} \text{ hält.} \end{array}$$

Eine partielle Funktion $f\colon \mathbb{N}^k \to \mathbb{N}$ heißt *berechenbar auf einer RM*, wenn eine RM R existiert, sodass f R-berechenbar ist.

Beachten Sie, dass die in Definition 5.8 definierte partielle Funktion undefiniert ist, wenn die RM R nicht hält. Für den Beweis des folgenden Satzes wird auf [6] verwiesen.

Satz 5.2

Jede partielle Funktion $f\colon \mathbb{N}^k \to \mathbb{N}$, die berechenbar auf einer RM ist, ist auf einer TM berechenbar und umgekehrt. □

5.4. Turingreduktion

Unter *Turingreduktion* oder einfach *Reduktion* versteht man die Reduktion eines Problems, dargestellt als formale Sprache, auf ein anderes Problem, sodass die entsprechende Abbildung berechenbar ist. Formal machen wir die folgende Definition.

Definition 5.9

Angenommen sind L, M Sprachen über dem Alphabet Σ und $R\colon \Sigma^* \to \Sigma^*$ eine Stringfunktion über Σ. Wir nennen R eine *Reduktion* wenn R berechenbar auf einer Turingmaschine ist und gilt

$$x \in L \Leftrightarrow R(x) \in M\ .$$

In diesem Fall schreiben wir auch $L \leqslant_m M$ und sagen, dass L auf M reduziert werden kann.

Wir verwenden Reduktionen, um die Schwierigkeiten bestimmter Probleme zu vergleichen, bzw. die Unentscheidbarkeit eines Problems zu zeigen. Dazu genügt es eine Reduktion von einer unentscheidbaren Sprache L auf die betrachtete Sprache M durchzuführen. Wäre das Problem M entscheidbar, könnten wir den Entscheidungsalgorithmus dafür dann ja für das Problem L anwenden. Somit erhalten wir einen Widerspruch zur Annahme, dass M entscheidbar ist, wie das folgende Lemma zeigt, siehe Aufgabe 5.10.

Lemma 5.1

Wenn $L \leqslant_m M$ und M rekursiv, dann ist L rekursiv.

Satz 5.3

Sei Σ ein Alphabet und $L \subseteq \Sigma^*$ rekursiv; dann ist auch das Komplement von L, $\sim L$ rekursiv.

Beweis. Da L rekursiv ist, gibt es eine totale TM M mit $L = \mathsf{L}(M)$. Wir definieren eine TM M', wobei der akzeptierende und der verwerfende Zustand von M vertauscht werden. Weil M total ist, ist auch M' total. Somit akzeptiert M' ein Wort genau dann, wenn M es verwirft und es folgt $\sim L = \mathsf{L}(M')$, d.h. $\sim L$ ist rekursiv. □

Satz 5.4

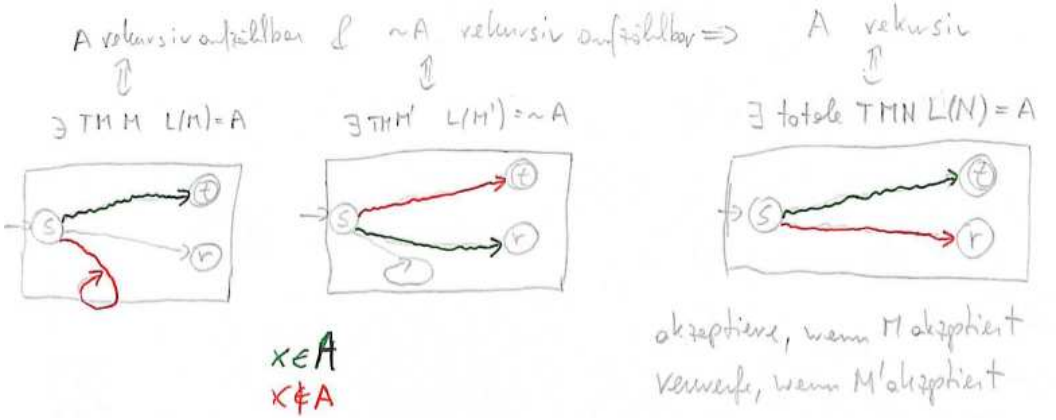
Jede rekursive Menge ist rekursiv aufzählbar. Andererseits ist nicht jede rekursiv aufzählbare Menge rekursiv.

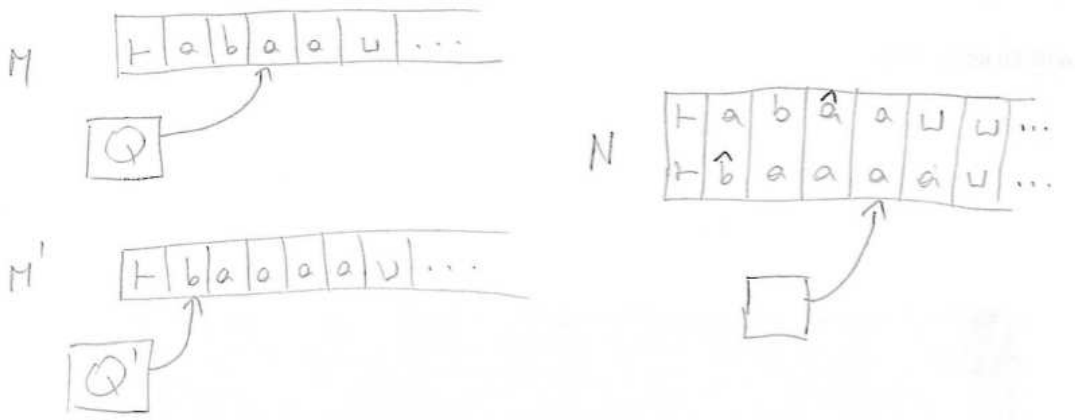
Beweis. Der erste Teil des Satzes ist eine Konsequenz der Definitionen; der zweite Teil wird in „Diskrete Strukturen“ bewiesen werden. □

Satz 5.5

Wenn L und $\sim L$ rekursiv aufzählbar sind, dann ist L rekursiv.

Beweis. Lösung.





Angenommen es existieren Turingmaschinen M und M', sodass $A = \mathsf{L}(M)$ und $\sim(A) = \mathsf{L}(M')$. Wir definieren eine neue TM N, die bei Eingabe x die Maschinen M und M' simultan auf x ausführt. Dazu teilen wir das Band von N in eine obere und untere Hälfte, sodass wir M auf der oberen und M' auf der unteren Hälfte ausführen können. Formal erreichen wir das, indem das Bandalphabet von N die folgenden Symbole enthält:

$$\boxed{\begin{matrix} a \\ c \end{matrix}} \qquad \boxed{\begin{matrix} \hat{a} \\ c \end{matrix}} \qquad \boxed{\begin{matrix} a \\ \hat{c} \end{matrix}} \qquad \boxed{\begin{matrix} \hat{a} \\ \hat{c} \end{matrix}}$$

Hier ist a ein Bandsymbol von M und c ein Bandsymbol von M'. Die Zusatzmarkierung $\hat{\cdot}$ wird verwendet, um die Position des Schreib-/Lesekopfes von M bzw. M' zu markieren. Das Band von N kann also folgende Gestalt haben:

$$\left|\begin{matrix} b \\ c \end{matrix}\right| \left|\begin{matrix} \hat{b} \\ c \end{matrix}\right| \left|\begin{matrix} a \\ c \end{matrix}\right| \left|\begin{matrix} b \\ d \end{matrix}\right| \left|\begin{matrix} a \\ d \end{matrix}\right| \left|\begin{matrix} a \\ d \end{matrix}\right| \left|\begin{matrix} a \\ c \end{matrix}\right| \left|\begin{matrix} a \\ \hat{c} \end{matrix}\right| \left|\begin{matrix} b \\ d \end{matrix}\right| \left|\begin{matrix} a \\ c \end{matrix}\right| \left|\begin{matrix} a \\ d \end{matrix}\right| \left|\begin{matrix} a \\ c \end{matrix}\right| \rangle \ldots$$

Die Maschine N führt abwechselnd einen Schritt von M oder einen Schritt von M' aus. Dazu werden die jeweiligen Zustände von M und M' durch zusätzliche Zustände in N gespeichert. Wenn M jemals akzeptiert, hält N und akzeptiert. Andererseits wenn M' akzeptiert, dann verwirft N. Nach Voraussetzung muss einer der Fälle für jede Eingabe auf jeden Fall eintreten. Somit hält N auf allen Eingaben und ist total. □

Im Beweis von Satz 5.5 haben wir bereits gesehen, dass das Band einer Turingmaschine geteilt werden kann. Im Folgenden untersuchen wir die Erweiterung von Turingmaschinen auf mehrere Bänder und Schreib-/Leseköpfe, wie in Abbildung 5.3 dargestellt. Allgemein hat die Übergangsfunktion einer k-bändigen Turingmaschine den folgenden Typ:

$$\delta \colon Q \times \Gamma^k \to Q \times \Gamma^k \times \{\mathrm{L}, \mathrm{R}\}^k .$$

Die formale Definition einer *k-bändigen deterministischen Turingmaschine* überlassen wir als Übung (Aufgabe 5.11).

Satz 5.6

Sei M eine k-bändige TM. Dann existiert eine (einbändige) TM M', sodass $\mathsf{L}(M) = \mathsf{L}(M')$.

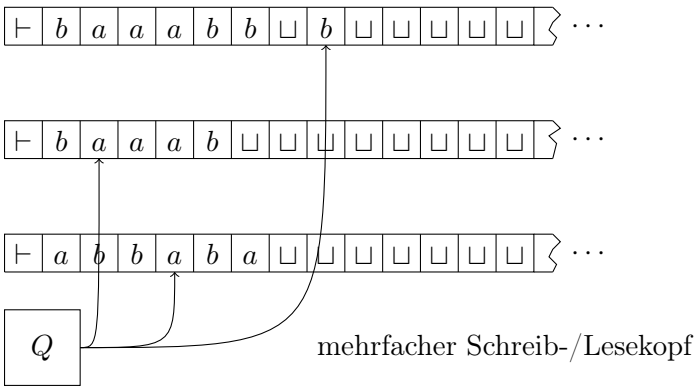


Abbildung 5.3.: 3-bändige Turingmaschine

Beweis. [Skizze] In der Literatur finden sich zwei unterschiedliche Ansätze, eine mehrbändige Turingmaschine durch eine einbändige zu simulieren. Entweder werden die k Bänder von M hintereinander, durch Sonderzeichen getrennt, auf dem Band von M' simuliert oder das Bandalphabet von M' besteht aus k-Tupeln von Bandsymbolen in M (mit möglichen Markierungen für die aktuelle Position des Schreib-/Lesekopfes). Dabei folgt man der Konstruktion im Beweis von Satz 5.5. Die formale Ausarbeitung dieser Beweisskizze überlassen wir als Übung (Aufgabe 5.12). □

5.5. Zusammenfassung

Alan Turing (1912–1954) schlug 1936 die Turingmaschine als allgemeines Berechnungsmodell vor. Das Ziel war auf möglichst intuitive Weise zu klären, was eigentlich einen „Algorithmus“ oder eine Berechnung ausmacht. In den 1930er Jahren war dies eine Frage, die eine Reihe von hochkarätigen Forschern wie etwa Alonzo Church (1903–1995), Kurt Gödel (1906–1978), Stephen Kleene (1909–1994) oder John von Neumann (1903–1957) zu lösen versuchten. Turings Ziel war es die Grenze zwischen dem was ein Computer berechnen kann und dem was er nicht berechnen kann, genau zu beschreiben. Seine Schlussfolgerungen treffen nicht nur auf seine abstrakten Turingmaschinen zu, sondern auch auf heutige, real existierende Computer oder Computersysteme. Beispielsweise seien Grenzen der Berechenbarkeit genannt, also Entscheidungsprobleme, die rein prinzipiell von einer Maschine nicht gelöst werden können.

Interessant ist, dass diese Untersuchungen über die Schranken der Berechenbarkeit zu einer Zeit entwickelt wurden in der die ersten Vorfahren moderner Rechner noch gar nicht gebaut waren. Die von Konrad Zuse (1910–1995) entwickelte Z3 wurde 1941 fertig gestellt und der von John Atanasoff (1903–1995) und Clifford Berry (1918–1963) entwickelte Atanasoff-Berry-Computer (ABC) wird auf 1939 datiert.

5.6. Aufgaben

Übungsaufgabe 5.1

Wie unterscheiden sich ein deterministischer endlicher Automat und eine Turingmaschine?

Lösung.

Eigenschaft	DEA	TM
Speicher	endlich (Zustände)	unendlich (Zustände und Band)
Eingabe	nur lesen	lesen und überschreiben
Eingabe	von links nach rechts	vor und zurück
akzept. Zustand	kann verlassen werden	kann nicht verlassen werden
akzept. Zustand	mehrere	genau einer
Eingabe x	nach $\|x\|$ Schritten abgearbeitet	muss nicht terminieren

Übungsaufgabe 5.2

Untersuchen Sie für die nachfolgenden Listen, ob es ein $m > 0$ und Indizes $i_1, i_2, \ldots, i_m$ gibt, sodass $x_{i_1} x_{i_2} \ldots x_{i_m} = y_{i_1} y_{i_2} \ldots y_{i_m}$.

1.

x_1	x_2	x_3	y_1	y_2	y_3
000	1	11	00	11	001

2.

x_1	x_2	x_3	y_1	y_2	y_3
000	001	010	00	100	101

Lösung.

1. 1132 ist eine Lösung, da $000\,000\,11\,1 = 00\,00\,001\,11$.
2. Es kann keine solchen Indizes geben. Wir betrachten drei Fälle:
   - Wenn $i_1 = 1$, dann ist das Wort $x_{i_1} \ldots x_{i_m}$ immer kürzer als das Wort $y_{i_1} \ldots y_{i_m}$.
   - Wenn $i_1 = 2$, dann unterscheiden sich die Wörter $x_{i_1} \ldots x_{i_m}$ und $y_{i_1} \ldots y_{i_m}$ (bereits am ersten Buchstaben).
   - Wenn $i_1 = 3$, dann unterscheiden sich die Wörter $x_{i_1} \ldots x_{i_m}$ und $y_{i_1} \ldots y_{i_m}$ (bereits am ersten Buchstaben).

Übungsaufgabe 5.3

Gegeben sei folgende Instanz des Postschen Korrespondenzproblems:
Untersuchen Sie für die nachfolgenden Listen, ob es ein $m > 0$ und Indizes $i_1, i_2, \ldots i_m$ gibt, sodass

$$x_{i_1} x_{i_2} \ldots x_{i_m} = y_{i_1} y_{i_2} \ldots y_{i_m} \ .$$

1.

x_1	x_2	x_3	y_1	y_2	y_3
11	00	1	00	0	11

2.

x_1	x_2	x_3	y_1	y_2	y_3
00	11	10	0	01	11

Übungsaufgabe 5.4

Seien $G = (V, \Sigma, R, S)$ eine Grammatik, $A \in V$, $u, v \in (V \cup \Sigma)^*$ und $x \in \Sigma^*$.
Welche der folgenden Aussagen ist falsch?

1. Wenn $A \overset{*}{\underset{G}{\Rightarrow}} x$, dann auch $uAv \overset{*}{\underset{G}{\Rightarrow}} uxv$.
2. Wenn $uAv \overset{*}{\underset{G}{\Rightarrow}} uxv$, dann auch $A \overset{*}{\underset{G}{\Rightarrow}} x$.
3. Wenn $S \overset{*}{\underset{G}{\Rightarrow}} x$, dann $x \in \mathsf{L}(G)$.
4. Eine der anderen Aussagen ist falsch.

Lösung.

1. ✓
2. ✗; die Eigenschaft gilt nur für kontextfreie Grammatiken.
3. ✓; $x \in \Sigma$ laut Angabe.
4. ✓

Übungsaufgabe 5.5

Betrachten Sie die Turingmaschine $M = (\{s, p, t, r\}, \{0, 1\}, \{\vdash, \sqcup, 0, 1\}, \vdash, \sqcup, \delta, s, t, r)$, wobei die (relevante Information der) Übergangsfunktion δ durch die *Zustandstabelle* angegeben ist:

	$\vdash$	0	1	$\sqcup$
s	$(s, \vdash, \mathrm{R})$	$(s, 0, \mathrm{R})$	$(s, 1, \mathrm{R})$	$(p, \sqcup, \mathrm{L})$
p	$(t, \vdash, \mathrm{R})$	$(t, 1, \mathrm{L})$	$(p, 0, \mathrm{L})$	·

1. Berechnen Sie die Schrittfunktion $\overset{1}{\underset{M}{\longrightarrow}}$ ausgehend von der Startkonfiguration $(s, \vdash 0011\sqcup^\infty, 0)$.
2. Welche Eingaben akzeptiert M, welche verwirft M?
3. Auf welchen Eingaben hält M, auf welchen nicht?
4. Beschreiben Sie $\mathsf{L}(M)$ in Worten.

Lösung.

2. M akzeptiert alle Eingaben.
3. M hält auf allen Eingaben.
4. $\mathsf{L}(M)$ ist die Menge aller Wörter über $\{0, 1\}$.

Übungsaufgabe 5.6

Sei $L = \{0^n 1^n \mid n \geqslant 0\}$.

1. Geben Sie eine Turingmaschine M an, welche die Sprache L akzeptiert. Dabei soll M folgenden Algorithmus verwenden.
   a) Wenn die Eingabe leer ist (also $\ldots \vdash \sqcup \ldots$), akzeptiere.
   b) Wenn das erste Zeichen der Eingabe eine Null ist, überschreibe sie mit $\vdash$. Ansonsten verwerfe.
   c) Gehe zum Ende der Eingabe (also zum ersten $\sqcup$).
   d) Wenn das letzte Zeichen der Eingabe eine Eins ist, überschreibe sie mit $\sqcup$. Ansonsten verwerfe.
   e) Wechsle zum Beginn der Eingabe und gehe zu Schritt i.
2. Testen Sie M auf den Eingaben ϵ, 0, 1, 01, 0011, 0101.

Lösung. Wir wählen $M = (\{s, t, r, q_0, q_1, q_2\}, \{0, 1\}, \{\vdash, 0, 1, \sqcup\}, \vdash, \sqcup, \delta, s, t, r)$ mit δ

	$\vdash$	0	1	$\sqcup$
s	$(s, \vdash, R)$	$(q_0, \vdash, R)$	$(r, \sqcup, R)$	$(t, \sqcup, R)$
q_0	$\cdot$	$(q_0, 0, R)$	$(q_0, 1, R)$	$(q_1, \sqcup, L)$
q_1	$(r, \vdash, R)$	$(r, \sqcup, R)$	$(q_2, \sqcup, L)$	$\cdot$
q_2	$(s, \vdash, R)$	$(q_2, 0, L)$	$(q_2, 1, L)$	$\cdot$

Dabei zeigt $\cdot$ einen beliebigen Übergang an (diese Situationen werden nicht erreicht. Die Übergänge für die Zustände t und r sind ebenfalls irrelevant, da diese Zustände nie mehr verlassen werden können.

Übungsaufgabe 5.7

Betrachten Sie die Turingmaschine M welche im folgenden Simulator definiert ist: http://morphett.info/turing/turing.html?417d522b074cb9937dadab228c042540. Seien Zustände $Q = \{a, b, c, \mathsf{halt}\}$, Eingabealphabet $\Sigma = \{\mathsf{X}\}$, Bandalphabet $\Gamma = \{\sqcup, \mathsf{X}\}$ und die folgende Übergangsfunktion δ:

$p \in Q$	$v \in \Sigma$	$\delta(p, v)$	$p \in Q$	$v \in \Sigma$	$\delta(p, v)$
a	$\sqcup$	(b, X, R)	a	X	$(\mathsf{halt}, \mathsf{X}, R)$
b	$\sqcup$	$(c, \sqcup, R)$	b	X	(b, X, R)
c	$\sqcup$	(c, X, L)	c	X	(a, X, L)

Zu Beginn sei das gesamte Band in beide Richtungen leer, d.h. mit dem Blanksymbol $\sqcup$ gefüllt und M starte im Zustand a. Wenn Sie den Simulator ausführen, werden Sie feststellen, dass M insgesamt sechs X auf das Band schreibt und dann hält.[a]
Fügen Sie nun einen weiteren Zustand d hinzu und modifizieren Sie δ entsprechend, um eine Turingmaschine zu konstruieren die mindestens zehn X auf das Band schreibt und hält. Verifizieren Sie Ihre Konstruktion mithilfe des Simulators.

Hinweis: Abweichend von der Definition in der Vorlesung verfügt M über ein *beidseitig* unendliches Band. Anstelle eines akzeptierenden und verwerfenden Zustandes gibt es nur den haltenden Zustand halt, da wir uns in diesem Beispiel nicht die Frage stellen ob M Eingaben akzeptiert oder verwirft, sondern das Verhalten von M auf einem anfänglich leeren Band untersuchen.

[a] Die Turingmaschine M ist eine Instanz eines „Busy Beavers“ mit drei Zuständen, siehe https://en.wikipedia.org/w/index.php?title=Busy_beaver&oldid=931091184.

Übungsaufgabe 5.8

Geben Sie ein Programm P für eine Registermaschine $R = ((x_i)_{1 \leqslant i \leqslant n}, P)$ an, welches eine Zuweisung $x_i := x_j$ durchführt.
Hinweis: Verwenden Sie Hilfsregister, falls nötig. Beachten Sie, dass der Wert von x_j vor sowie nach der Zuweisung der selbe sein soll.

Lösung. Wir benötigen ein Hilfsregister x_k und verfahren in drei Schritten:

1. Setzte Register x_i und x_k auf Null.

```
while x_i ≠ 0 do
    x_i := x_i − 1
end;
while x_k ≠ 0 do
    x_k := x_k − 1
end
```

2. Kopiere Inhalt von x_j nach x_i und x_k (danach hat Register x_j den Inhalt Null).

```
while x_j ≠ 0 do
    x_i := x_i + 1;
    x_k := x_k + 1;
    x_j := x_j − 1
end
```

3. Setze x_j wieder auf den alten Wert (mithilfe von x_k).

```
while x_k ≠ 0 do
    x_j := x_j + 1;
    x_k := x_k − 1
end
```

Übungsaufgabe 5.9

Geben Sie ein Programm P für eine Registermaschine $R = ((x_i)_{1 \leqslant i \leqslant 5}, \mathsf{P})$ an, sodass R bei Eingabe $(m, n, 0, 0, 0)$ mit Registerinhalt $(m, n, m + n, 0, 0)$ hält.

Hinweis: Benutzen Sie die Register x_4 und x_5 als Hilfsregister, falls nötig.

Lösung.

```
while x_1 ≠ 0 do
    x_1 := x_1 − 1;
    x_3 := x_3 + 1;
    x_4 := x_4 + 1
end;
while x_2 ≠ 0 do
    x_2 := x_2 − 1;
    x_3 := x_3 + 1;
    x_5 := x_5 + 1
end;
while x_4 ≠ 0 do
    x_4 := x_4 − 1;
    x_1 := x_1 + 1
end;
while x_5 ≠ 0 do
    x_5 := x_5 − 1;
    x_2 := x_2 + 1
end
```

Übungsaufgabe 5.10

Beweisen Sie, dass wenn L, M formale Sprachen über Σ, sodass $L \leqslant_m M$ und M rekursiv, dann ist L rekursiv.

Übungsaufgabe 5.11

Geben Sie eine formale Definition für eine k-bändige deterministische Turingmaschine an. Wie sieht die Startkonfiguration aus? Wie ist die akzeptierte Sprache definiert?

Übungsaufgabe 5.12

Beweisen Sie Satz 5.6.

6.
Einführung in die Komplexitätstheorie

In diesem kurzen Kapitel werden die grundlegenden Begriffe der Komplexitätstheorie eingeführt, wie etwa die Komplexitätsklassen P und NP. Außerdem wird die Komplexitätstheorie kurz motiviert. Wir beschränken uns dabei auf Zeitkomplexitätsklassen und wollen vor allem den Zusammenhang zur Berechenbarkeitstheorie (siehe Kapitel 5) betonen.

Die Komplexitätstheorie analysiert Algorithmen und Probleme in Bezug auf die verwendeten Ressourcen. Es geht hier also um Fragen wie etwa nach der notwendigen Rechenzeit eines bestimmter Algorithmus oder allgemeiner eines bestimmtes Problems.

Während sich die *Analyse von Algorithmen* vor allem mit Fragen zur Effizienz eines bestimmten Algorithmus beschäftigt, versucht die *Komplexitätstheorie* allgemeinere Aussagen über die Schwierigkeit eines *Problems* zu treffen. Wir sagen zum Beispiel SAT ist in NP, um auszudrücken, dass das Problem der Erfüllbarkeit einer aussagenlogischen Formeln zwar nicht leicht (also in polynomieller Zeit) gelöst werden kann, dass aber eine Lösung leicht (also in polynomieller Zeit) verifiziert werden kann.

Probleme werden in der Komplexitätstheorie als formale Sprachen charakterisiert, wie wir auch gleich in der Folge an einem Beispiel sehen werden.

6.1. Laufzeitkomplexität

Sei M eine *totale* TM, dann ist die *Laufzeitkomplexität* von M eine Funktion $T\colon \mathbb{N} \to \mathbb{N}$, wobei T wie folgt definiert ist

$$T(n) := \max\{m \mid M \text{ hält bei Eingabe } x,\ |x| = n, \text{ nach } m \text{ Schritten}\} .$$

$T(n)$ bezeichnet die *Laufzeit* von M, wenn n die Länge der Eingabe von M bezeichnet. In diesem Fall nennen wir M eine *T-Zeit-Turingmaschine*.

Definition 6.1

Sei $T\colon \mathbb{N} \to \mathbb{N}$ eine numerische Funktion:

$$\mathsf{DTIME}(T) := \{\mathsf{L}(M) \mid M \text{ ist eine mehrbändige TM mit Laufzeit (ungefähr) in } T\} .$$

Wobei „ungefähr“ korrekt bedeutet, dass die Laufzeit von M *asymptotisch* durch T gebunden ist, das heißt es existiert eine Konstante $c \in \mathbb{R}^+$ und eine untere Schranke $m \in \mathbb{N}$, sodass für alle $n \geqslant m$, die Laufzeit von M bei Eingabe $x \leqslant c{\cdot}T(n)$, wobei $|x| = n$. Die Klasse $\mathsf{DTIME}(T)$ enthält also alle Sprachen, deren Zugehörigkeitstest von einer mehrbändigen TM in Zeit T entschieden werden kann.

Beispiel 6.1

Für die TM M aus Beispiel 5.1 ergibt sich z.B. $T(0) = 3$ und $T(1) = 5$. Weiters sehen wir, dass $T(n)$ asymptotisch linear in n ist.

Lösung.

- $(s, \vdash \sqcup^\infty, 0) \xrightarrow[M]{1} (s, \vdash \sqcup^\infty, 1) \xrightarrow[M]{1} (p, \vdash \sqcup^\infty, 0) \xrightarrow[M]{1} (t, \vdash \sqcup^\infty, 1)$
- $(s, \vdash 0\sqcup^\infty, 0) \xrightarrow[M]{1} (s, \vdash 0\sqcup^\infty, 1) \xrightarrow[M]{1} (s, \vdash 0\sqcup^\infty, 2) \xrightarrow[M]{1} (p, \vdash 0\sqcup^\infty, 1) \xrightarrow[M]{1} (t, \vdash 1\sqcup^\infty, 0)$
- $(s, \vdash 1\sqcup^\infty, 0) \xrightarrow[M]{1} (s, \vdash 1\sqcup^\infty, 1) \xrightarrow[M]{1} (s, \vdash 1\sqcup^\infty, 2) \xrightarrow[M]{1} (p, \vdash 1\sqcup^\infty, 1) \xrightarrow[M]{1} (p, \vdash 0\sqcup^\infty, 0) \xrightarrow[M]{1} (t, \vdash 0\sqcup^\infty, 1)$

Wir interessieren uns für das asymptotische Wachstum von Komplexitätsschranken. Dabei sind Laufzeitkomplexitätsfunktionen, die polynomiell sind, solchen die exponentiell wachsen (im Normalfall) deutlich vorzuziehen.

Im Besonderen interessieren wir uns für diejenigen Probleme, die durch einen Algorithmus, der in polynomieller Zeit läuft, berechnet werden können. Diese Probleme werden in der Komplexitätsklasse P zusammengefasst. Hier steht das P für *P*olynomielle Zeit. Die Klasse P umfasst jene Probleme, für welche eine Lösung in polynomieller Zeit *gefunden* werden kann.

Definition 6.2

$$\mathsf{P} := \bigcup_{k \geqslant 1} \mathsf{DTIME}(n^k)$$

Beispiel 6.2

Betrachte SAT als Sprache:

$$\text{SAT} = \{F \mid F \text{ Formel mit erfüllbarer Belegung } \mathsf{v}\} .$$

Dann gilt $\text{SAT} \in \mathsf{DTIME}(2^n)$, aber es ist nicht bekannt ob $\text{SAT} \in \mathsf{P}$.

Wie wir im Beispiel sehen ist es nicht unbedingt einfach festzustellen, ob eine aussagenlogische Formel F erfüllbar ist oder nicht. Wir kennen zwar sehr viele Möglichkeiten die Erfüllbarkeit zu überprüfen und haben davon auch in der Vorlesung bereits einige kennengelernt (siehe Kapitel 2). Allerdings brauchen alle diese Methoden im schlimmst möglichen Fall exponentiell viele Schritte, um die Erfüllbarkeit zu entscheiden. Andererseits ist es sehr einfach eine gegebene Belegung v daraufhin zu kontrollieren, ob sie die Erfüllbarkeit von F zeigt.

Das bedeutet, dass das Problem SAT schwer zu lösen, aber eine vorgeschlagene Lösung leicht zu kontrollieren ist. Diese Eigenschaft haben sehr viele Probleme in der Informatik und wir können diese in eine eigene Komplexitätsklasse—nämlich NP—zusammenfassen. Die nächste Definition formalisiert die Intuition, dass eine vorgeschlagene Lösung eines gegebenen Problems leicht zu kontrollieren ist. Beachten Sie dazu, dass das Zertifikat c als mögliche Lösung des Problems interpretiert werden kann.

Definition 6.3: Polytime Verifikator

Ein *Verifikator* einer Sprache $L \subseteq \Sigma^*$, ist ein Algorithmus V sodass

$$L = \{x \in \Sigma^* \mid \exists\ c, \text{sodass } V \text{ akzeptiert Eingabe } (x, c)\}\ .$$

Ein *polytime Verifikator* ist ein Verifikator mit (asymptotischer) Laufzeit n^k wobei $|x| = n$. Das Wort c wird *Zertifikat* genannt.

Definition 6.4

NP ist die Klasse der Sprachen, die einen polytime Verifikator haben.

Alternativ kann die Klasse NP auch mit Hilfe von *nichtdeterministischen* Turingmaschinen definiert werden. Diese, historisch frühere Methode erklärt auch den Namen der Klasse; NP steht für *N*ichtdeterministische *P*olynomielle Zeit. Da NP aber genau jene Probleme umfasst, für welche eine Lösung in polynomieller Zeit *verifiziert* werden kann, hat sich mittlerweile die hier gegebene Definition eingebürgert.

6.2. Reduktionen in polynomieller Zeit

Basierend auf der Turingreduzierbarkeit (siehe Kapitel 5) definieren wir Reduktionen, die in polynomieller Zeit berechenbar sind.

Sie M eine totale TM mit k Bändern und Eingabealphabet Σ. Angenommen M läuft in polynomieller Zeit und schreibt bei Eingabe $x \in \Sigma^*$, $R(x)$ auf das (erste) Band. Dann nennen wir $R\colon \Sigma^* \to \Sigma^*$ *in polynomieller Zeit berechenbar.*

Definition 6.5: Reduzierbarkeit

Seien L und M formale Sprachen über dem Alphabet Σ. Sei $R\colon \Sigma^* \to \Sigma^*$ in polynomieller Zeit berechenbar, sodass gilt

$$x \in L \Leftrightarrow R(x) \in M\ .$$

Dann ist L *in polynomieller Zeit* auf M *reduzierbar*; kurz: $L \leqslant^{\mathrm{p}} M$.

Beispiel 6.3

Seien

$$L = \{x \in \{a, b\}^* \mid |x| \text{ ist gerade}\} \text{ und}$$
$$M = \{x \in \{a, b\}^* \mid x \text{ ist ein Palindrom gerader Länge}\}\ .$$

Dann gilt $L \leqslant^{\mathrm{p}} M$.

Lösung. Wir geben eine in polynomieller Zeit berechenbare Abbildung $R\colon \{a, b\}^* \to \{a, b\}^*$ an, sodass $x \in L \Leftrightarrow R(x) \in M$. Wir wählen $R(a) = a$ und $R(b) = a$. Somit wird ein Wort aus $\{a, b\}^n$ in a^n umgewandelt. Genau dann wenn n gerade ist, ist a^n ein Palindrom gerader Länge. Tabellarisch ergibt sich:

$x \in L$	x	$R(x)$	$R(x) \in M$
✓	ϵ	ϵ	✓
×	a	a	×
×	b	a	×
✓	aa	aa	✓
✓	ab	aa	✓
✓	ba	aa	✓
✓	bb	aa	✓
×	aaa	aaa	×
⋮	⋮	⋮	⋮

Satz 6.1

Seien L und M formale Sprachen mit $L \leqslant^{\mathrm{p}} M$. Dann gilt

- Wenn $M \in \mathsf{P}$, dann $L \in \mathsf{P}$.
- Wenn $M \in \mathsf{NP}$, dann $L \in \mathsf{NP}$.

Beweis. Direkte Konsequenz aus Definition 6.5. □

Definition 6.6: $\mathcal{C}$-hart

Sei $\mathcal{C}$ eine beliebige Komplexitätsklasse und L eine Sprache über Σ. Angenommen für alle Sprachen $M \in \mathcal{C}$ gilt: $M \leqslant^{\mathrm{p}} L$. Dann sagen wir L ist *$\leqslant^{p}$-hart für $\mathcal{C}$* oder kurz L ist *$\mathcal{C}$-hart*.

Beispiel 6.4

SAT ist $\leqslant^{\mathrm{p}}$-hart für NP.

Definition 6.7: $\mathcal{C}$-vollständig

Für eine Sprache L, gelte

1. L $\leqslant^{\mathrm{p}}$-hart für $\mathcal{C}$ und
2. $L \in \mathcal{C}$.

Dann ist L *$\leqslant^{p}$-vollständig für $\mathcal{C}$* oder (kurz) *$\mathcal{C}$-vollständig*.

Bemerkung

Laut Definition 6.7 ist jedes nicht trivale Problem in P auch P-vollständig (bezüglich $\leqslant^{\mathrm{p}}$), weil man das Problem in der Reduktion lösen kann. Ein Problem L ist *trivial*, wenn $L = \varnothing$ oder $L = \Sigma^*$.

Den folgenden Satz geben wir ohne Beweis.

Satz 6.2

SAT ist NP-vollständig. □

Als einfache Konsequenz des obigen Satzes sehen wir, dass $\mathsf{P} = \mathsf{NP}$, wenn $\mathrm{SAT} \in \mathsf{P}$. Somit könnten Sie mit einem deterministischen Algorithmus für SAT, der in polynomieller Zeit läuft, eine Million Dollar verdienen.

6.3. Zusammenfassung

Im Jahr 1969 führte Stephen Cook (1939–) Turings Untersuchung mit der Frage fort welche Probleme *effektiv*, also mit vertretbarem Aufwand, algorithmisch zu lösen sind. Die Klasse der manchmal als *nicht handhabbar* (*intractable*) betrachteten Probleme werden als NP-hart bezeichnet. Diese Begriffsbestimmung geht auf Cook zurück. Für diese Arbeiten ist Cook 1982 mit dem *Turing Award* ausgezeichnet worden.

Es ist sehr unwahrscheinlich, dass die exponentielle Steigerung der Rechengeschwindigkeit, die bei der Computerhardware erzielt worden ist („Mooresches Gesetz“), sich bemerkenswert auf unsere Fähigkeit auswirken wird, umfangreiche Beispiele solcher nicht handhabbaren Probleme berechnen zu können. Allerdings haben die Forschungen und Erkenntnisse der letzten Jahre gezeigt, dass NP-harte Probleme keineswegs „nicht handhabbar“ sind. Die Frage, ob eine gegebene aussagenlogische Formel erfüllbar ist, wäre ein solches Problem. In den letzten Jahren (bzw. Jahrzehnten) wurden sehr mächtige automatische Techniken entwickelt, um dieses Problem (fast immer) effizient lösen zu können [12].

6.4. Aufgaben

Übungsaufgabe 6.1

Bestimmen Sie die Laufzeitkomplexität der Turingmaschinen aus den Beispielen 5.1 und 5.3.

Hinweis: Wenn Sie die Funktion $T(n)$ nicht genau angeben können, bestimmen Sie das Wachstum ungefähr, also asymptotisch.

Lösung. Zuerst betrachten wir die TM M aus Beispiel 5.1. Die Eingabe der Länge n, auf der M die meisten Schritte ausführt ist 1^n. Dann gilt

$$(s, \vdash 1^n \sqcup^\infty, 0) \xrightarrow[M]{2n+3} (t, \vdash 1^n \sqcup^\infty, 1)$$

und somit ist $T(n) = 2n + 3$, also ist T linear.

Für die TM aus Beispiel 5.3 sind die Eingaben, welche eine maximale Anzahl an Schritten erlauben Palindrome. Betrachten wir z.B. 0^n. Die erste Schleife der TM läuft in linearer Zeit, da $(s, \vdash 0^n \sqcup^\infty, 0) \xrightarrow[M]{2n+1} (s, \vdash\vdash 0^{n-2} \sqcup^\infty, 1)$ Die nächste Schleife ist kürzer, nämlich $2(n-2)+1$, weil das erste und letzte Zeichen gelöscht wurden, etc. Insgesamt sind maximal $\frac{n}{2}$ dieser Schleifen möglich und somit ergibt sich die (ungefähre) Laufzeitkomplexität n^2.

Übungsaufgabe 6.2

Skizzieren Sie eine zweibändige TM M mit linearer Laufzeitkomplexität, deren Sprache die Menge aller Palindrome ungerader Länge über dem Alphabet $\{0, 1\}$ ist. Warum wurde die Einschränkung auf Palindrome ungerader Länge getroffen?

Übungsaufgabe 6.3

Betrachten Sie das *Rucksackproblem*. Es sind n verschiedene Gegenstände mit einem bestimmten Gewicht g_i und Wert w_i $(1 \leqslant i \leqslant n)$ gegeben. Aus diesen Gegenständen soll eine Auswahl getroffen werden, die in einen Rucksack mitgenommen werden können.
Der Rucksack darf maximal ein Gewicht von G bekommen (andernfalls würden wir unter dem Gewicht zusammenbrechen). Gleichzeitig sollen zumindest Gegenstände in einem Wert von W eingepackt werden (andernfalls wäre der Inhalt des Rucksacks nutzlos). Formal drücken wir das wie folgt aus. Das Problem ist gelöst, wenn es eine Indexmenge $I \subseteq \{i \mid 1 \leqslant i \leqslant n\}$ gibt, sodass gilt:

$$\sum_{j \in I} g_j \leqslant G$$
$$\sum_{j \in I} w_j \geqslant W$$

1. Geben Sie (i) eine lösbare und (ii) eine unlösbare Instanz des Problems an, indem Sie die Parameter entsprechend setzen.
2. Bezeichne $((g_i)_{1 \leqslant i \leqslant n}, (w_i)_{1 \leqslant i \leqslant n}, G, W)$ eine bestimmte Instanz des Rucksackproblems. Geben Sie an welche Information (welches *Zertifikat*) Sie benötigen, um schnell (also in polynomieller Zeit) entscheiden zu können, dass das Problem lösbar ist.
3. Das Problem ist NP-vollständig. Geben Sie in Pseudocode einen Algorithmus an, der für eine fixe Instanz und ein geeignetes Zertifikat in polynomieller Zeit überprüft, dass das Problem lösbar ist. Sie sollen also einen polytime Verifikator beschreiben.

Übungsaufgabe 6.4

Seien A, B und C formale Sprachen. Zeigen Sie, dass die Reduktion mit polynomieller Zeit transitiv ist:

Wenn $A \leqslant^{\mathrm{p}} B$ und $B \leqslant^{\mathrm{p}} C$, dann $A \leqslant^{\mathrm{p}} C$.

Übungsaufgabe 6.5

Seien A und B formale Sprachen und $\mathbb{C}$ eine Komplexitätsklasse. Zeigen Sie:

Wenn A $\mathbb{C}$-hart ist und $A \leqslant^{\mathrm{p}} B$, dann ist B $\mathbb{C}$-hart.

Gilt auch

$$\text{Wenn } A \text{ } \mathbb{C}\text{-vollständig ist und } A \leqslant^{\mathrm{p}} B, \text{ dann ist } B \text{ } \mathbb{C}\text{-vollständig.}$$

Hinweis: Verwenden Sie das Resultat von Aufgabe 6.4.

Übungsaufgabe 6.6

Gegeben ein (Un-)Gleichungssystem $A\vec{x} \geq \vec{b}$, wobei $A \in \mathbb{Z}^{m \times n}$, $\vec{x}$ ein n-Vektor von Unbekannten und b ein m-Vektor mit Einträgen aus $\mathbb{Z}$. Das finden einer Lösung $\vec{x}$ nennt man *integer programming*.
Zeigen Sie, dass integer programming NP-vollständig ist.
Hinweis: Reduzieren Sie von SAT, dh. geben Sie eine in polynomieller Zeit berechenbare Funktion r an, sodass

- $r(\varphi) = A\vec{x} \geq \vec{b}$ und
- φ erfüllbar genau dann, wenn $A\vec{x} \geq b$ eine Lösung besitzt.

Sie können annehmen, dass die aussagenlogische Formel in KNF gegeben ist.

7.
Einführung in die Programmverifikation

In diesem kurzen Kapitel werden wir auf die Prinzipien der Analyse von Programmen eingehen und die Begriffe *Verifikation* und *Validierung* von Software klären (Abschnitt 7.1). Darüber hinaus werden wir die Verifikation nach Hoare in Abschnitt 7.2 behandeln. Schließlich diskutieren wir in Abschnitt 7.3 den historischen Kontext und in Abschnitt 7.4 (optionale) Aufgaben, die zur weiteren Vertiefung dienen sollen.

7.1. Prinzipien der Analyse von Programmen

Grundsätzlich besteht der Wunsch nach fehlerfreier Software. Die Praxis zeigt jedoch, dass dies bisher als nicht realisierbar angesehen werden muss. So ist bei einem größeren Softwareprojekt, wie etwa der Entwicklung eines Betriebssystems mit mehreren Millionen Zeilen Code, auch davon auszugehen, dass wiederum Millionen von Fehlern enthalten sein werden. *Verifikation* dient nun zum Nachweis, dass die Spezifikation eines Programms auch korrekt implementiert wurde. Nur wenn so sichergestellt ist, dass die Implementierung nicht von der Spezifikation abweicht, wenn das Programm also verifiziert ist, können wir davon ausgehen, dass das Programm wirklich fehlerfrei ist. Eine unvollständige Methode der Verifikation stellt das Testen von Software dar. Sehr viele Fehler lassen sich durch Testen finden, aber das Testen von Programmen kann nur die Fehler erkennen, auf die auch tatsächlich getestet wird. Hingegen versuchen formale Methoden der Verifikation, die Korrektheit eines Programms zu zeigen. Die Verifikation muss von der *Validierung* abgegrenzt werden. Letztere überprüft nicht, ob das Programm die Spezifikation korrekt implementiert, sondern ob die Spezifikation unseren Anforderungen entspricht. Kurz gefasst kann man sagen, dass die Verifikation überprüft, ob wir die Dinge richtig tun, wohingegen die Validierung überprüft, ob wir das Richtige tun.

Heutzutage werden in der Verifikation immer häufiger formale Methoden verwendet. Diese Methoden erlauben es die Verifikation frühzeitig in den Entwicklungsprozess einzubinden. Das ist wichtig, da Fehler umso leichter ausgebessert werden können, je früher diese erkannt werden. Formale Methoden erlauben es auch die Verifikationstechniken effizient zu gestalten. Das kann bis zur Automatisierung der Verifikation führen. So kann erreicht werden, dass die Verifikation eines Programms zeitlich abgekürzt wird. Formale Methoden werden mittlerweile auch verwendet, um geeignete Teststrategien beziehungsweise Testmengen zu finden.

Im nächsten Abschnitt werden wir die Verifikation nach Hoare näher betrachten. Es ist wichtig darauf hinzuweisen, dass die Verifikation nach Hoare nicht vollständig automatisiert werden kann, zumindest wenn der Quellcode übliche Programmkonstrukte, wie etwa Prozeduren verwendet. Es ist in modernen Softwarepaketen undenkbar, den Code monolithisch zu entwerfen. Somit stößt die Anwendbarkeit dieser Methode schnell an ihre Grenzen. Deshalb spielt der Hoare-Kalkül in der Praxis eine untergeordnete Rolle. Andererseits bietet der Hoare-Kalkül eine gute Einführung in das Gebiet der Verifikation und die entsprechenden Arbeiten von Hoare gehören zu den meistzitierten Arbeiten in der Informatik.

Ein anderes Beispiel einer formalen Methode, die zur Verifikation verwendet werden kann, ist das *Model Checking*. Model Checking ist eine formale Methode, die anhand einer endlichen Beschreibung (Modell) eines Systems (oder Programs), systematisch prüft, ob das beschriebene System eine bestimmte formale Eigenschaft besitzt. Model Checking ist eine automatische Methode [4].

Eine wichtige Begriffsabgrenzung in der Verifikation ist der Unterschied zwischen *totaler* und *partieller* Korrektheit. Totale Korrektheit schließt die Terminierung des untersuchten Programms mit ein. Partielle Korrektheit überprüft zwar, dass der Quelltext korrekte Ergebnisse liefert, aber geht von der Terminierung der untersuchten Programme aus.

7.2. Verifikation nach Hoare

Um interessante Eigenschaften von Programmen ausdrücken zu können, reichen die Möglichkeiten der Aussagenlogik nicht aus. Wir müssen unseren logischen Formalismus erweitern. Die Grundlage der Verifikation nach Hoare sind *prädikatenlogische* Ausdrücke. Eine Einführung in die Prädikatenlogik geht über diese Vorlesung hinaus. Wir werden uns also hier auf eine einfache Teilklasse der Prädikatenlogik beschränken: Wir betrachten *atomare Formeln* und deren Abschluss unter aussagenlogischen Junktoren.[1] Für die vollständige Definition von prädikatenlogischen Formeln wird auf die Vorlesung „Logic in Computer Science“ beziehungsweise auf [10, 5] verwiesen. Die wichtigste Erweiterung sind so genannte *Prädikatensymbole*, die wir anhand eines einfachen Beispiels motivieren.

Beispiel 7.1

Angenommen wir haben die Konstante $\mathsf{7}$ und das Prädikatensymbol $\mathsf{ist\_prim}$ in unserer Sprache, dann können wir $\mathsf{ist\_prim}(\mathsf{7})$ schreiben, um auszudrücken, dass 7 eine Primzahl ist.

Prädikatensymbole erlauben es uns über Elemente einer Menge, im Beispiel die natürlichen Zahlen, Aussagen zu treffen. Um Elemente in dieser Menge zu referenzieren, verwendet man *Terme*.

Definition 7.1: Signatur

Eine *Signatur* F ist eine Menge von *Funktionssymbolen*, sodass jedem Symbol $f \in \mathsf{F}$ eine *Stelligkeit* n zugeordnet wird. Symbole mit Stelligkeit 0 werden auch *Konstanten* genannt.

Definition 7.2: Term

Sei F eine Signatur und sei V eine (unendliche) Menge von *Variablen*, sodass $\mathsf{F} \cap \mathsf{V} = \varnothing$. Die Menge $\mathsf{T}(\mathsf{F}, \mathsf{V})$ aller *Terme (über* F*)* ist induktiv definiert:

1. Jedes Element von V ist ein Term.
2. Wenn $n \in \mathbb{N}$ und $f \in \mathsf{F}$ mit Stelligkeit n sowie $t_1, \ldots, t_n$ Terme sind, dann ist auch $f(t_1, \ldots, t_n)$ ein Term.

[1] Dies ist eine sehr starke Vereinfachung. In der vollständigen Definition werden den hier betrachteten aussagenlogischen Junktoren auch noch Quantoren ($\forall$, $\exists$) als logische Symbole zur Seite gestellt [10, 5]. Erst dann kann man eigentlich von prädikatenlogischen Formeln sprechen.

Als besonderes Prädikatensymbol, betrachten wir die *Gleichheit*. Das Gleichheitszeichen können wir als Prädikatensymbol auffassen, denn die Gleichung drückt eine Beziehung zwischen den Objekten aus, auf welche die Terme s und t verweisen. Wegen seiner Bedeutung wird die Gleichheit hier aber besonders hervorgehoben. Eine Gleichung $s = t$ drückt somit aus, dass die beiden Terme s und t als gleich angesehen werden.

Sei nun P ein Prädikatensymbol und $t_1, \ldots, t_n$ Terme über einer geeignet gewählten Signatur. Der Ausdruck $P(t_1, \ldots, t_n)$ sowie die Gleichung $t_1 = t_2$ wird *Atom* oder *atomare Formel* genannt. Mit Hilfe von atomaren Formeln als Basis können nun *Zusicherungen* definiert werden.

Definition 7.3: Zusicherungen

Wir definieren *Zusicherungen* induktiv.

1. Atome sind Zusicherungen.
2. Wenn A und B Zusicherungen sind, dann sind $\neg A$, $(A \wedge B)$, $(A \vee B)$ und $(A \to B)$ auch Zusicherungen.

Der Einfachheit halber nennen wir Zusicherungen manchmal auch *Formeln*, da keine Verwechslungsgefahr mit aussagenlogischen Formeln besteht.

Zusicherungen bilden den logischen Formalismus, der es uns erlaubt, den Zustand eines Programms zu beschreiben. Zusicherungen sind rein syntaktisch definiert, das heißt wir müssen diesen Formeln eine Bedeutung zuordnen. Dazu verwendet man *Interpretationen*, die wir als Verallgemeinerungen von Algebren verstehen können. Eine Interpretation $\mathcal{I}$ gibt an, wie die Symbole einer Formel zu verstehen sind. Etwa würden wir in Beispiel 7.1 das Symbol $\mathsf{ist\_prim}$ so interpretieren wollen, dass das Atom $\mathsf{ist\_prim}(n)$ wahr wird gdw. n eine Primzahl ist. In ähnlicher Weise werden Gleichungen $t_1 = t_2$ über einer gegeben Interpretation $\mathcal{I}$ wahr genannt werden gdw. die Terme t_1 und t_2 in $\mathcal{I}$ als gleich angesehen werden.

Ist die Wahrheit von Atomen in $\mathcal{I}$ definiert, ist es leicht, die Wahrheit einer beliebigen Zusicherung über $\mathcal{I}$ zu definieren: Wie in Kapitel 2 verwenden wir die Wahrheitstabellen für die aussagenlogischen Junktoren $\neg$, $\wedge$, $\vee$ und $\to$, um die Wahrheit einer zusammengesetzten Formel zu überprüfen. Wenn eine Formel F in einer Interpretation $\mathcal{I}$ wahr ist, schreibt man $\mathcal{I} \models F$. Schließlich wollen wir noch ausdrücken können, dass eine bestimmte Zusicherung B aus einer Prämisse A folgt. Dazu definiert man $A \models B$ gdw. für alle Interpretationen $\mathcal{I}$, sodass $\mathcal{I} \models A$ auch $\mathcal{I} \models B$ gilt. Die Relation $\models$ wird *Konsequenzrelation* genannt und erweitert die in Definition 2.3 eingeführte Konsequenzrelation für aussagenlogische Formeln. Wir motivieren die Konsequenzrelation anhand des nächsten Beispiels.

Beispiel 7.2

Seien $x < 5$ und $x + 1 < 7$ Atome und damit Zusicherungen, wobei wir „5“ und „7“ durch die Zahlen 5 und 7 interpretieren und auch die Relation „<“ durch die natürliche Ordnung auf den natürlichen Zahlen. Dann gilt intuitiv die folgende Folgerung für alle natürlichen Zahlen, die wir für x einsetzen können:

$$x < 5 \models x + 1 < 7 \,.$$

Konvention

Wenn nicht anders angegeben, werden die Symbole der Arithmetik (wie $<$, $\leqslant$ sowie natürliche Zahlen) in der üblichen Art und Weise interpretiert. Diese Annahme verwenden wir in der Folge immer implizit, wenn wir über Konsequenzen sprechen.

Bevor wir nun endlich die Regeln des Hoare-Kalkül definieren können, benötigen wir noch die Definition einer *Substitution*.

Definition 7.4: Substitution

Sei F eine Signatur und V eine Menge von Variablen. Eine *Substitution* ist eine Abbildung $\sigma\colon \mathsf{V} \to \mathsf{T}(\mathsf{F},\mathsf{V})$, sodass $\sigma(x) \neq x$ für höchstens endlich viele Variablen x. Die (möglicherweise leere) endliche Menge der Variablen, die durch die Abbildung σ nicht auf sich selber abgebildet werden, nennt man den *Definitionsbereich* $\mathsf{dom}(\sigma)$ von σ. Wenn $\mathsf{dom}(\sigma) = \{x_1, \ldots, x_n\}$ können wir σ wie folgt schreiben:

$$\sigma = \{x_1 \mapsto \sigma(x_1), \ldots, x_n \mapsto \sigma(x_n)\} \ .$$

Man sagt oft dass eine Variable x durch eine Substitution σ *instanziiert* wird, wenn gilt $x \in \mathsf{dom}(\sigma)$.

Definition 7.5

Jede Substitution σ kann zu einer Abbildung auf Termen $\overline{\sigma}\colon \mathsf{T}(\mathsf{F},\mathsf{V}) \to \mathsf{T}(\mathsf{F},\mathsf{V})$ erweitert werden:

$$\overline{\sigma}(t) := \begin{cases} \sigma(t) & \text{wenn } t \in \mathsf{V} \\ f(\overline{\sigma}(t_1), \ldots, \overline{\sigma}(t_n)) & \text{wenn } t = f(t_1, \ldots, t_n) \ . \end{cases}$$

Die Anwendung (der Erweiterung) einer Substitution auf einen Term ersetzt simultan alle Variablen im Definitionsbereich durch ihr Bild.

Wenn es nicht zu Verwechslungen kommen kann, dann bezeichnen wir die Erweiterung $\overline{\sigma}$ einer Substitution σ, wiederum mit σ.

In Definition 7.4 haben wir Substitutionen für Terme eingeführt, es ist aber intuitiv einleuchtend, wie eine Substitution $\{x \mapsto t\}$ auf eine Formel F angewandt werden soll: Wenn x in F vorkommt, werden alle Vorkommnisse von x durch den Term t ersetzt. Für diesen Prozess schreiben wir einfach $F\{x \mapsto t\}$. In Korrektheitsbeweisen werden Ausdrücke, die in Programmen auftreten, etwa die rechten Seiten von Zuweisungen, als Terme repräsentiert.

Mit diesem Handwerkszeug wenden wir uns der Verifikation nach Hoare zu. Wir nehmen die in Definition 5.7 definierten while-Programme als Grundlage, um den Hoare Kalkül erklären zu können. Die hier betrachteten Hoare-Regeln bilden einen Teil der üblicherweise in der Literatur betrachteten Regeln, da sie der Einfachheit halber auf while-Programme spezialisiert sind (siehe Kapitel 5).

Definition 7.6

Sei P ein while-Programm. Ein *Hoare-Tripel* ist wie folgt definiert:

$$\{Q\}\ P\ \{R\}\ ,$$

$$[\text{z}] \quad \frac{}{\{Q\{x \mapsto t\}\}\ x := t\ \{Q\}} \qquad [\text{a}] \quad \frac{\{Q'\}\ P\ \{R'\}}{\{Q\}\ P\ \{R\}}\ Q \models Q', R' \models R$$

$$[\text{s}] \quad \frac{\{Q\}\ P_1\ \{R\} \quad \{R\}\ P_2\ \{S\}}{\{Q\}\ P_1; P_2\ \{S\}} \qquad [\text{w}] \quad \frac{\{I \wedge B\}\ P\ \{I\}}{\{I\}\ \mathsf{while}\ B\ \mathsf{do}\ P\ \mathsf{end}\ \{I \wedge \neg B\}}$$

Abbildung 7.1.: Regeln nach Tony Hoare

wobei Q und R Zusicherungen sind. Dabei wird Q *Vorbedingung* und R wird *Nachbedingung* genannt.

Definition 7.7: Hoare-Kalkül

Seien Q, R Zusicherungen und P ein while-Programm. Dann heißt P *korrekt in Bezug auf* Q *und* R wenn

$$\{Q\}\ P\ \{R\}\ ,$$

mit Hilfe der in Abbildung 7.1 dargestellten Regeln abgeleitet werden kann. Die Hoare-Regeln werden üblicherweise von unten nach oben gelesen: Das Problem die Korrektheit des Programms zu zeigen, wird sukzessive in kleinere Teile aufgespaltet.
Wir nennen P *partiell korrekt* für eine Spezifikation S, wenn P korrekt ist in Bezug auf Zusicherungen Q und R die der Spezifikation S entsprechen.

Intuitiv drückt das Hoare Triple $\{Q\}\ P\ \{R\}$ aus, dass wenn *vor* der Ausführung des Programms P die Aussage Q hält, dann *nach* der Ausführung die Aussage R gilt. Hier setzen wir Termination von P voraus.

Wir motivieren die in Abbildung 7.1 dargestellten Regeln kurz. Die erste Regel [z], ist ein Axiom des Kalküls, da keine Vorbedingungen erfüllt sein müssen. Das Axiom wird verwendet, um *Zuweisungsbefehle* korrekt in den Zusicherungen abbilden zu können. Die Regel [a] dient dazu, die Vorbedingung abzuschwächen (Nebenbedingung $Q \models Q'$) beziehungsweise die Nachbedingung zu verstärken (Nebenbedingung $R' \models R$). Von oben nach unten gelesen wird also die Aussage *abgeschwächt*, deshalb die Bezeichung [a]. Um diese Regel anwenden zu können, müssen die Nebenbedingungen erfüllt sein. Die Regel [s] dient dazu die Korrektheit der einzelnen Teile eines Programms separat zu beweisen. Hier wird also die *Sequentialität* des Programms ausgenutzt. Die vierte Regel [w] wird auch while-Regel genannt, da sie die Korrektheit von while-Schleifen überprüft. Wir bezeichnen die Formel I als *Schleifeninvariante* oder kurz *Invariante*.

Definition 7.8

Angenommen P ist ein partiell korrektes Programm für eine gegebene Spezifikation. Wenn es noch gelingt die Terminierung von P nachzuweisen, dann ist auch der Beweis der *totalen Korrektheit* gelungen.

7.3. Zusammenfassung

Schon in den Anfängen der Programmierung wurde erkannt, dass es wünschenswert wäre, die Korrektheit eines Programms formal verifizieren zu können und sich nur auf Fehlersuche beschränken zu müssen. Etwa hatte sich schon Alan Turing dieser Fragestellung gewidmet. Robert Floyd (1963–2001) hat diese Ideen weitergeführt. Auf die Arbeiten von Floyd aufbauend, hat Charles Antony Richard (oder Tony) Hoare (1934–), den hier vorgestellten Hoare-Kalkül entwickelt. Hoare wurde 1980 der *Turing Award* für seine Arbeiten zu Definition und Design von Programmiersprachen verliehen.

Die Erkenntnis, dass der Hoare-Kalkül für komplexere Programmkonstrukte ungeeignet ist, geht auf Edmund M. Clarke (1945–) zurück, der gemeinsam mit Allen Emerson (1954–) das Model Checking als formale Verifikationsmethode von endlichen Systemen vorgeschlagen hat. Clarke, Emerson und Joseph Sifakis (1946–) wurden gemeinsam für ihre Arbeiten zu Model Checking im Jahr 2007 mit dem *Turing Award* ausgezeichnet.

7.4. Aufgaben

Übungsaufgabe 7.1

Verwenden Sie die *natürliche Interpretation*. Welche Konsequenzen gelten?

1. $x_1 + x_2 = m + n \models x_1 = m \wedge x_2 = n$
2. $x_1 + x_2 = m + n \models x_1 = m \vee x_2 = n$
3. $x_1 = m \wedge x_2 = n \models x_1 + x_2 = m + n$
4. $x_1 \geqslant 0 \wedge x_2 \geqslant 0 \models x_1 + x_2 \geqslant 0$
5. $x_1 \geqslant x_2 \wedge x_1 \geqslant n \models x_2 \geqslant n$
6. $x_1 \geqslant 0 \wedge x_2 = 0 \models x_1 + x_2 \geqslant 0$

Lösung.

1. ✗
2. ✗
3. ✓
4. ✓
5. ✗
6. ✓

Übungsaufgabe 7.2

Betrachten Sie das while-Programm P

$$\begin{array}{l}\textsf{while } x_1 \neq 0 \textsf{ do}\\ \quad x_1 := x_1 - 1;\\ \quad x_2 := x_2 + 1\\ \textsf{end}\end{array}$$

1. Leiten Sie das Hoare-Tripel

$$\{x_1 = m \wedge x_2 = n\}\ P\ \{x_2 = m + n\}$$

im Hoare-Kalkül ab.
Hinweis: Verwenden Sie die Schleifeninvariante $x_1 + x_2 = m + n$.

2. Ist P partiell korrekt bzw. total korrekt?

Übungsaufgabe 7.3

Untersuchen Sie für jede der nachfolgenden Aussagen ob sie wahr/falsch ist.

1. Eine Instanz des Postschen Korrespondenzproblems (PCP) hat entweder keine Lösung oder unendlich viele.
2. Es gibt ein While-Programm $P_1; P_2$, sodass P_2 niemals ausgeführt wird.
3. Es gibt eine Turingmaschine M und ein Wort x, sodass M die Eingabe x sowohl akzeptiert als auch verwirft.

Lösung.

1. Wahr. Wenn es keine Lösung gibt, ist die Behauptung gezeigt. Wenn es eine Lösung $i_1, \ldots, i_m$ gibt, dann ist auch $i_1, \ldots, i_m, i_1, \ldots, i_m$, etc. eine Lösung.
2. Ja, z.B.

$$\begin{array}{l}x_1 := x_1 + 1;\\ \textsf{while } x_1 \neq 0 \textsf{ do}\\ \quad x_2 := x_2 + 1;\\ \textsf{end}; x_1 := x_1 + 1\end{array}$$

3. Falsch, da M deterministisch.

Teil IV.

Appendix

Anhang A.

Appendix

A.1. Logische Schaltkreise

In diesem Abschnitt wenden wir Boolesche Algebren auf *logische Schaltkreise* (oft auch *Schaltnetze* genannt) an. Ein logischer Schaltkreis ist eine abstrakte Repräsentation eines *elektronischen Schaltkreises* der eine Funktion implementiert, die zum Beispiel als Eingabewert eine hohe/niedere Spannung erhält und als Ausgangswert eine hohe/niedere Spannung zurückliefert. Elektronische Schaltkreise bilden die Grundlage der Informationsverarbeitung in heutigen Rechnerarchitekturen. Wir beschäftigen uns hier nicht mit der tatsächlichen *Realisierung* von logischen Schaltkreisen, dieses Gebiet wird in der technischen Informatik untersucht, sondern betrachten logische Schaltkreise abstrakt als besondere Instanz einer Booleschen Algebra.

Definition A.1: Logischer Schaltkreis

Sei $\mathbb{B} = \{0, 1\}$, wobei 0 als niedere Spannung und 1 als hohe Spannung interpretiert wird. Wir betrachten die Algebra

$$\langle \mathbb{B}; +, \cdot, \sim, 0, 1 \rangle ,$$

wobei die Operationen $+$, $\cdot$, $\sim$ wie in Abbildung 3.1 definiert sind. Diese Boolesche Algebra heißt *Schaltalgebra*. Ein *logischer Schaltkreis* ist ein algebraischer Ausdruck der Schaltalgebra, wobei die Operationen $+$, $\cdot$, $\sim$ als *logische Gatter* dargestellt werden. Diese Gatter heißen das *Oder-*, das *Und-* und das *Nicht-Gatter*. Die drei Gatter sind in Abbildung A.1 dargestellt.

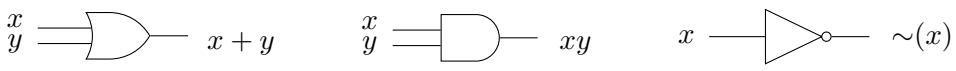


Abbildung A.1.: Logische Gatter

Alternativ zu Definition A.1 können wir die betrachteten Gatter als Repräsentationen von logischen Junktoren $\vee$, $\wedge$ und $\neg$ betrachten und die Spannungswerte 0 und 1 als F beziehungsweise T interpretieren. Analog zu Definition 3.13 erweitern wir die Schaltalgebra zu einer Algebra von *Schaltfunktionen*.

Definition A.2

Seien n, m natürliche Zahlen. Sei Abb die Menge der Abbildungen von $\mathbb{B}^n$ nach $\mathbb{B}^m$. Wir betrachten die Boolesche Algebra

$$\langle \mathsf{Abb}; +, \cdot, \sim, \mathbf{0}, \mathbf{1} \rangle ,$$

wobei $\mathbf{0}$ und $\mathbf{1}$ konstante Funktionen bezeichnen und $+$, $\cdot$, $\sim$ punktweise definiert sind.

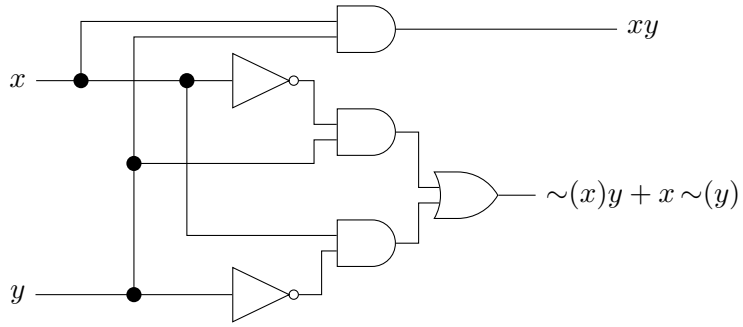


Abbildung A.2.: Halbaddierer

Diese Boolesche Algebra wird als *Algebra der n-stelligen Schaltfunktionen bezeichnet.*

Wie schon bei Booleschen Ausdrücken entspricht jeder logische Schaltkreis einer Schaltfunktion und umgekehrt. Außerdem gilt Satz 3.1 analog für Schaltfunktionen.

Satz A.1

Seien F, G logische Schaltkreise (in den Variablen $x_1, \ldots, x_n$) und seien $f\colon \mathbb{B}^n \to \mathbb{B}$, $g\colon \mathbb{B}^n \to \mathbb{B}$ ihre Schaltfunktionen. Dann gilt $F \approx G$ gdw. $f = g$ in der Algebra der Schaltfunktionen.

Da logische Schaltkreise nur eine andere Darstellung von Booleschen Ausdrücken sind, folgt, dass jede Aussage über die Gleichheit von Schaltfunktionen eine Aussage über die Äquivalenz von Booleschen Ausdrücken ist und somit für *alle* Booleschen Algebren gilt (und umgekehrt).

Logische Schaltkreise können in vielfältiger Weise kombiniert werden, um einfache arithmetische Operationen, etwa binäre Addition, zu implementieren. Wir realisieren die Funktionen „Übertrag“ und „Summand“ und es ist leicht einzusehen wie aus dem so erhaltenen *Halbaddierer* die binäre Addition zu implementieren ist [7].

Der „Übertrag“ $\mathsf{carry}(a, b)$ ist 1 gdw. $a = 1$ und $b = 1$. Also können wir carry mit Hilfe einer Konjunktion darstellen: $\mathsf{carry}(a, b) = ab$. Nun betrachten wir die „Summand“-Funktion $\mathsf{summand}(a, b)$. Hier erkennen wir, dass $\mathsf{summand}(a, b) = 1$ gdw. $a = 0$ und $b = 1$ gilt oder $a = 1$ und $b = 0$. Somit erhalten wir die folgende Schaltfunktion: $\mathsf{summand}(a, b) = \sim(a)b + a\sim(b)$. Wenn wir die Operationen carry und $\mathsf{summand}$ in geeigneter Weise kombinieren, erhalten wir einen sogenannten *Halbaddierer*. Eine mögliche Realisierung ist in Abbildung A.2 dargestellt, wo x und y Boolesche Variablen bezeichnen.

Allerdings ist der so erhaltene Schaltkreis zur Realisierung eines Halbaddierers nicht optimal. Wir verwenden zwei Nicht-Gatter, drei Und-Gatter und ein Oder-Gatter. Eine Anwendung der

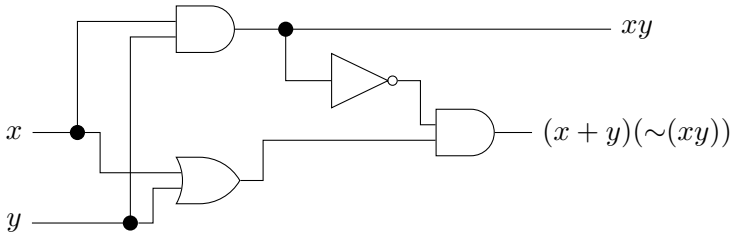


Abbildung A.3.: Einfacherer Halbaddierer

Gesetze der Schaltalgebra zeigt, dass wir den Halbaddierer auch mit 4 Gattern realisieren können.

$$
\begin{aligned}
\sim(a)b + a\sim(b) &= (\sim(a)b + a)(\sim(a)b + \sim(b)) && \text{Definition 3.7}\\
&= (a + \sim(a)b)(\sim(b) + b\sim(a)) && +, \cdot \text{ kommutativ}\\
&= (a + \sim(a)b)(\sim(b) + \sim(\sim(b))\sim(a)) && \text{Lemma 3.11}\\
&= (a + b)(\sim(b) + \sim(a)) && \text{Lemma 3.9}\\
&= (a + b)(\sim(a) + \sim(b)) && + \text{ kommutativ}\\
&= (a + b)\sim(ab) && \text{Lemma 3.12 .}
\end{aligned}
$$

Die vereinfachte Realisierung des Halbaddierers ist in Abbildung A.3 dargestellt.

Dieses Beispiel zeigt, dass es möglich ist die selbe Schaltfunktion mit einfacheren Schaltkreisen zu realisieren, wenn die entsprechenden Booleschen Ausdrücke vereinfacht werden. Üblicherweise wird die Größe eines Schaltkreises über die Anzahl der notwendigen Komponenten einer logisch äquivalenten DNF oder KNF gemessen.

Definition A.3

Eine *minimale DNF* D eines Booleschen Ausdruckes E ist eine DNF von A, sodass die Anzahl der Konjunktionen in D minimal ist. Wenn zwei DNFs von E die gleiche Anzahl von Konjunktionen haben, dann ist jene DNF minimal, deren Anzahl von Literalen minimal ist. Die *minimale KNF* ist analog definiert.

Aus Satz 3.3 folgt, dass eine minimale DNF beziehungsweise minimale KNF immer existiert. Allerdings ist es nicht immer leicht die minimale DNF oder KNF zu finden. Auf die dazu entwickelten Verfahren, wie etwa das Erstellen von Karnaugh-Veitch-Diagrammen gehen wir hier nicht näher ein, sondern verweisen auf die Vorlesung „Rechnerarchitektur“ beziehungweise die Literatur zu Technischen Informatik [8].

Übungsaufgabe A.1

Vereinfachen Sie das Schaltnetz in Abbildung A.1.

Hinweis: Wandeln Sie das Schaltnetz in einen Booleschen Ausdruck um, vereinfachen Sie diesen soweit wie möglich, und zeichnen Sie das Ergebnis als neues Schaltnetz.

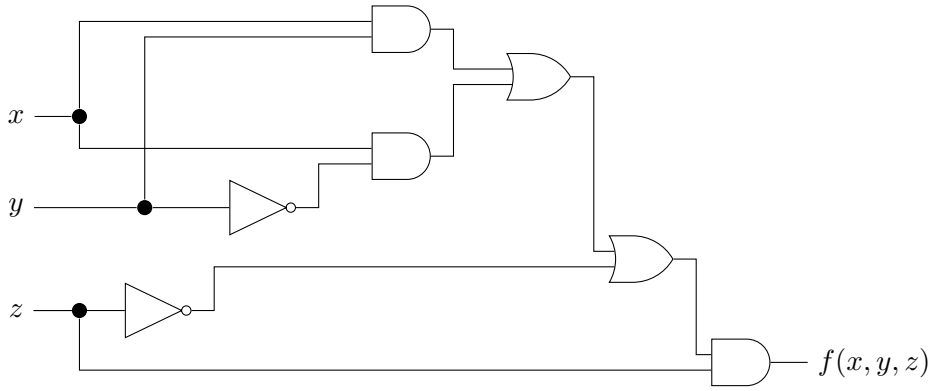


Abbildung A.4.: Schaltnetz zu Aufgabe A.1.

Lösung. Wir lesen aus dem Schaltnetz folgenden Booleschen Ausdruck ab und vereinfachen:

$$\begin{aligned}&(((xy) + (x \sim(y))) + \sim(z))z\\ &= (x + \sim(z))z\\ &= (xz) + (\sim(z)z)\\ &= (xz) + 0\\ &= xz\end{aligned}$$

Das resultierende Schaltnetz ignoriert den Eingang y.

A.2. Anwendung kontextfreier Grammatiken: XML

Die klassische Anwendung von kontextfreien Sprachen findet sich in *Parsergeneratoren* oder allgemeiner im Compilerbau. Ein Parsergenerator verwandelt die Beschreibung einer Sprache in einen Parser für diese Sprache. Die Sprache wird üblicherweise mit Hilfe einer kontextfreie Grammatik angegeben. Parser werden zur syntaktischen Analyse von Programmen verwendet. Neuere Anwendungen finden im Bereich der Wissensrepräsentation statt, etwa in *XML*-Dokumenten. Ein essentieller Teil von XML ist die DTD, die *Document Type Definition*, die im Prinzip eine kontextfreie Sprache ist. In der Folge gehen wir nur auf die modernere Anwendung ein.

XML steht für *eXtensible Markup Language* und ist eine *Markup-Sprache* oder *Kennzeichnungssprache* wie HTML. Im Gegensatz zu HTML, dessen Aufgabe die *Formatierung* des Textes ist, ist die Aufgabe von XML den *Inhalt* des Textes zu beschreiben. In XML haben wir die Möglichkeit, durch benutzerdefinierte Tags eine Aussage über den Text, der zwischen den Tags steht, zu machen. Angenommen wir wollen ausdrücken, dass ein bestimmter Teil des Textes einen Zeitungsartikel beschreiben soll. Dann führen wir das Tag `ARTICLE` ein und schreiben:

`<ARTICLE>` *Artikelbeschreibung* `</ARTICLE>`

Solche Tags werden *Namenselemente* genannt. Wie aber geben wir einem Namenselement Inhalt? Dazu werden entweder „*Document Type Definitions*“ (DTDs) oder *XML-Schemata* verwendet. Eine *DTD* hat die Form

`<!DOCTYPE` *Name der DTD* `[`
Liste der Elementbeschreibungen `]>`

Um Elementbeschreibungen definieren zu können, verwendet man (erweiterte) *reguläre Ausdrücke* [9]. Wir definieren diese induktiv, wobei wir die Bedeutung der Ausdrücke teilweise nur informell einführen.

1. – Namenselemente sind reguläre Ausdrücke
   – Der Ausdruck `#PCDATA`, der jedes Wort ohne XML-Tags bezeichnet ist ein regulärer Ausdruck.
2. – $E \mid F$ bezeichnet die *Vereinigung* der durch E und F beschriebenen Elemente,
   – E, F bezeichnet die *Konkatenation* der durch E und F beschriebenen Elemente,
   – $E\star$ ($E+$) steht für die beliebige (beliebige, aber mindestens einmalige) Wiederholung der durch E beschriebenen Elemente,
   – $E?$ steht für die Möglichkeit die durch E beschriebenen Elemente optional anzugeben.

Beispiel A.1

Wir betrachten die folgende *Document Type Definition*:

```
<!DOCTYPE NEWSPAPER [

<!ELEMENT NEWSPAPER (ARTICLE+)>
<!ELEMENT ARTICLE (HEADLINE,BYLINE,LEAD,BODY,NOTES)>
<!ELEMENT HEADLINE (#PCDATA)>
<!ELEMENT BYLINE (#PCDATA)>
<!ELEMENT LEAD (#PCDATA)>
<!ELEMENT BODY (#PCDATA)>
<!ELEMENT NOTES (#PCDATA)>

]>
```

Der *Name* der DTD ist `NEWSPAPER`. Das erste Element–dem Startsymbol einer Grammatik entsprechend–ist ebenfalls `NEWSPAPER`. Die Elementbeschreibung drückt aus, dass das Element `NEWSPAPER` eine nichtleere Sequenz von Artikeln beschreibt. Schließlich ist ein `ARTICLE` die Verknüpfung der folgenden Textelemente:

`HEADLINE`	die Kopfzeile
`BYLINE`	der Untertitel
`LEAD`	die Einleitung
`BODY`	der eigentliche Artikel
`NOTES`	Anmerkungen

In der Folge wandeln wir exemplarisch zwei Elementbeschreibungen in Produktionsregeln einer kontextfreien Grammatik um. Die Definition

```
<!ELEMENT ARTICLE (HEADLINE,BYLINE,LEAD,BODY,NOTES)>
```

entspricht der Regel

$$\texttt{ARTICLE} \rightarrow \texttt{HEADLINE BYLINE LEAD BODY NOTES}\,.$$

Nun betrachten wir die Zeile

```
<!ELEMENT NEWSPAPER (ARTICLE+)> ,
```

und wollen diese Beschreibung durch Regeln einer kontextfreien Grammatik ausdrücken. Wir müssen dazu die (leichte) Schwierigkeit bewältigen, dass in der Elementbeschreibung von einer Variante eines Kleene-Sterns Gebrauch gemacht wird. Diese Schwierigkeit bewältigen wir durch die Einführung eines zusätzlichen Nichtterminals `ARTICLES` und erhalten die folgenden drei Regeln:

$$\begin{aligned} \texttt{NEWSPAPER} &\rightarrow \texttt{ARTICLES} \\ \texttt{ARTICLES} &\rightarrow \texttt{ARTICLE} \mid \texttt{ARTICLE ARTICLES} \end{aligned}$$

Man kann allgemein zeigen, dass jede Regel mit (erweiterten) regulären Ausdrücken im Rumpf durch eine Sammlung äquivalenter gewöhnlicher Regeln ersetzt werden kann [9].

Literaturverzeichnis

[1] Alfred V. Aho, Monica Sin-Ling Ravi Sethi Lam, and Jeffrey D. Ullman. *Compilers: Principles, Techniques, and Tools.* Addison-Wesley series in computer science / World student series edition. Addison-Wesley, 1986.

[2] F. Baader and T. Nipkow. *Term Rewriting and All That.* Cambridge University Press, 1998.

[3] H. Barendregt. *The Lambda Calculus: Its Syntax and Semantics.* Studies in Logic and the Foundations of Mathematics. Elsevier, 2te auflage edition, 1985.

[4] E. M. Clarke, O. Grumberg, and D. A. Peled. *Model Checking.* MIT Press, 2000.

[5] H.-D. Ebbinghaus, J. Flum, and W. Thomas. *Einführung in die mathematische Logik.* Hochschul Taschenbuch. Spektrum Akademischer Verlag, 5te auflage edition, 2007.

[6] K. Erk and L. Priese. *Theoretische Informatik: Eine umfassende Einführung.* Springer Verlag, 3te auflage edition, 2008.

[7] J. L. Hein. *Discrete Structures, Logic, and Computability.* Jones and Bartlett Publishers, 3te auflage edition, 2010.

[8] D.W. Hoffmann. *Grundlagen der Technischen Informatik.* Hanser, 4te auflage edition, 2014.

[9] J. E. Hopcroft, R. Motwani, and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, 2001. 2te Auflage.

[10] M. Huth and M. Ryan. *Logic in Computer Science: Modelling and Reasoning about Systems.* Cambridge University Press, 2te auflage edition, 2004.

[11] D. Kozen. *Automata and Computability.* Springer Verlag, 1997.

[12] D. Kroening and O. Strichman. *Decision Procedures – An Algorithmic Point of View.* Springer Verlag, 2008.

[13] M. Schaper. Programming turing machines, 2011. Bachelor Thesis.

[14] S. Singh. *Fermats letzter Satz: Die abenteuerliche Geschichte eines mathematischen Rätsels.* Deutscher Taschenbuch Verlag, 2000.