

Automaten und Formale Sprachen

1. Vorlesung



Prof. Dr. Dietrich Kuske



FG Automaten und Logik, TU Ilmenau

Wintersemester 2024/25

Einführung

Grundbegriffe

Rechtslineare Sprachen

Kontextfreie Sprachen

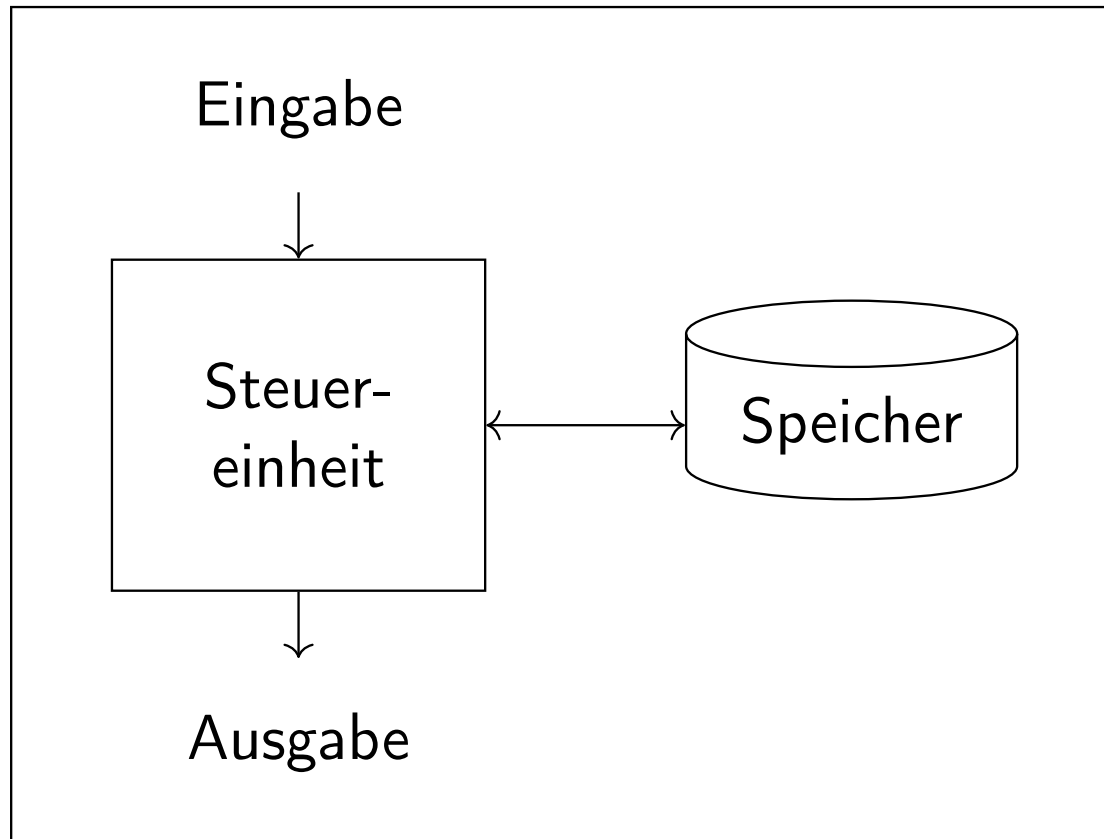
Grundthema von AFS und BuK

Grundfrage

Welche **Probleme** können von einem Algorithmus bzw. von einer **Maschine** mit beschränkten **Ressourcen** gelöst werden und welche können nicht gelöst werden?

- Was ist ein „Problem“?
- Was meinen wir mit „Maschine“?
- Welche „Ressourcen“ betrachten wir und wie werden diese „beschränkt“?

Maschinen



Steuereinheit:

- endlich viele Zustände
- ändern sich in Abhängigkeit von Eingabe und Speicherkonfiguration
- produziert Ausgaben
- gibt „Speicherbefehle“

Eingabe: Folge von „Buchstaben“

Ausgabe: gut/schlecht, Folge von „Buchstaben“ oder $\mathbb{N}$

Probleme

Wir betrachten Probleme als Abbildungen

$$\begin{array}{ccc} f: & \text{Menge der mgl. Eingaben} & \longrightarrow \text{Menge der mgl. Ausgaben} \\ & E & \longrightarrow A \end{array}$$

$E =$ alle „Wörter über Eingabealphabet“ oder Tupel natürlicher Zahlen

$A =$ gut/schlecht ($\hat{=}\{1, 0\}$), alle „Wörter über Ausgabealphabet“, oder natürliche Zahlen

Spezialfall $A = \{0, 1\}$

Solche Probleme f heißen **Entscheidungsprobleme**.

Sie sind gegeben durch die Menge $\{w \in E \mid f(w) = 1\} \subseteq E$. Solche Mengen nennen wir „Sprachen“.

beschränkte Ressourcen

- Art des Speicherzugriffs: verboten / Kellerspeicher / Register / Turing-Band
- Art der Steuereinheit: deterministisch / nichtdeterministisch
- Dauer der Berechnung: polynomiell / exponentiell / ...
- Größe des Speichers: fest / logarithmisch / polynomiell / ...

Ressourcen für AFS

- 5 Leistungspunkte ergeben **150 h** Arbeitsaufwand
(§4 PStO-AB Bachelor, Master und Diplom: „Ein Leistungspunkt entspricht ... einer Arbeitszeit von 30 Stunden.“)
- das Semester hat 15 Wochen, also **10 h / Woche**
- eine Vorlesung, $\frac{1}{2}$ Übung: $2\frac{1}{4}$ h / Woche Präsenzveranstaltungen
- es bleiben also

$7\frac{3}{4}$ h / Woche für das Selbststudium

= Vorlesungs- und Übungsvor- und -nachbereitung

Arbeitsweise

1. Sie kommen natürlich zu jeder Vorlesung und hören aktiv zu.
2. Der anspruchsvolle Stoff kann nicht durch alleiniges Hören verstanden werden. Daher werden Sie den Vorlesungsstoff semesterbegleitend nacharbeiten: Definitionen („Konzepte“) und Sätze („Sachverhalte“) herausschreiben und auswendig lernen, Beweise („Begründungen“) verstehen (= wiedergeben können), weitere Literatur zu Rate ziehen
3. Übungen leben von Ihrer Mitarbeit, sie werden (frühzeitig!) vor- und später nachgearbeitet.
4. Zu jeder Veranstaltung bringen Sie sämtliche Unterlagen zum Nachschlagen mit.
5. Bei Verständnisproblemen fragen Sie bitte frühzeitig!
6. Kein Vorlesungs- oder Übungsmaterial wird von Ihnen auf irgendwelchen Servern veröffentlicht.

Organisatorisches zur Vorlesung

Informationen, aktuelle Version der Folien, Übungsblätter und Mitschnitte der Vorlesungen: Website der Veranstaltung (Startpunkt: www.tu-ilmenau.de/a1, nicht moodle)

Damit wir Informationen an Sie schicken können, melden Sie sich bitte trotzdem im moodle-Kurs an

Prüfung: 90-minütige Klausur im Februar/März nächsten Jahres

Literaturempfehlung: Uwe Schöning: Theoretische Informatik – kurz gefasst, Spektrum Akademischer Verlag

Organisatorisches zur Übung

Die **Übungen** führt Herr M.Sc. Pfau durch.

Bonuspunkte 1: Die Lösungen der Übungsaufgaben werden abgegeben, (teilweise) korrigiert, und sie ergeben (anteilig) 7% der Klausurpunkte.

Abgabe der Übungsaufgaben: bis Freitag 11 Uhr (ungerade Woche) in Briefkasten vor Z 1047 oder in Freitagsübung

In den Übungen werden mögliche Lösungen und Probleme besprochen ...

Bonuspunkte 2: ... und kurze Tests geschrieben (in jeder 2. Übung, erstmalig am 8., 11. und 13.11.), in denen Definitionen, Sätze usw. abgefragt werden, diese ergeben (anteilig) 3% der Klausurpunkte.

Darüber hinaus führt Herr Pfau eine regelmäßige **Konsultation** durch, bei der Sie Fragen zum Vorstellungsstoff stellen können (Do U 11 Uhr, Raum Z 2073).

Einführung

Grundbegriffe

Rechtslineare Sprachen

Kontextfreie Sprachen

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

Definition

Sei X eine Menge.

- $X^* = \bigcup_{n \in \mathbb{N}} X^n$ ist die Menge der Tupel beliebiger Länge über X .
- $X^+ = \bigcup_{n \geq 1} X^n = X^* \setminus \{()\}$ ist die Menge der nichtleeren Tupel über X .

Beispiele

Elemente von $\mathbb{N}^*$: $(2, 3, 2, 2, 1, 6), (4, 3, 0, 1), (3, 0, 5), (3, 3), (1), ()$

Elemente von $\{a, b, c, d\}^*$: $(a, b, a), ()$

Elemente von $\{A, \dots, Z, 0, \dots, 9\}^*$: $(A, F, S, 2, 0, 2, 3), ()$

$$\emptyset^* = \{()\}$$

$$\emptyset^+ = \emptyset$$

Definition

Sei X Menge und $w = (x_1, \dots, x_n) \in X^*$.

- $|w| = n$ ist die **Länge** von w .
- Für $x \in X$ ist $|w|_x$ die Anzahl der Vorkommen von x in w , d.h.
 $|w|_x = |\{i \mid 1 \leq i \leq n, x_i = x\}|$.

Bemerkung

Es gelten also

- $X^n = \{w \in X^* \mid |w| = n\}$ und
- $|()| = 0$.
- $|(a, b, b, a, c)|_a = 2, |(a, b, b, a, c)|_c = 1$

Definition

Ein **Alphabet** ist eine endliche nichtleere Menge.

üblicherweise heißen unsere Alphabete Σ , Γ , Δ

Beispiele

- Alphabete: $\{0\}$, $\{0, 1, 2\}$, ...
 $\{A, \dots, Z, 0, \dots, 9\}$
 $\{\text{groß}, \text{klein}\}$
- keine Alphabete: $\emptyset$, $\mathbb{N}$, $\mathbb{Q}$

Ist Σ Alphabet, so nennen wir die Elemente oft **Buchstaben**.

Beispiel

Das Alphabet $\{0, 1, 2\}$ hat also die drei Buchstaben 0, 1 und 2.

Das Alphabet $\{\square, \diamond\}$ hat die zwei Buchstaben $\square$ und $\diamond$.

Ist Σ ein Alphabet, so heißen die Elemente von Σ^* auch **Wörter über Σ** (auch: **String**, **Zeichenkette**).

Beispiel

- (0) , $()$ und $(1, 2, 0, 0)$ sind also Wörter über dem Alphabet $\{0, 1, 2\}$.
- $(\square)$, $(\diamond, \square)$, $(\diamond, \square, \diamond)$ und $()$ sind Wörter über dem Alphabet $\{\square, \diamond\}$.

Schreibweise: Für das Wort $(a_1, a_2, \dots, a_n)$ schreiben wir auch $a_1 a_2 \dots a_n$, typische Wörter: $u, v, w, x, y, z, w_0, w_1, \dots$

Beispiel

$(1, 2, 0, 0)$ wird geschrieben als 1 2 0 0
 (1) wird geschrieben als 1
 $()$ wird geschrieben als ε (das **leere Wort**)

Definition

Sind $u = (a_1, a_2, a_3, \dots, a_m)$ und $v = (b_1, b_2, b_3, \dots, b_n)$ Wörter, so ist $u \cdot v$ das Wort $(a_1, a_2, a_3, \dots, a_m, b_1, b_2, b_3, \dots, b_n)$; es wird als **Verkettung** oder **Konkatenation** von u und v bezeichnet.

Schreibweise: An Stelle von $u \cdot v$ schreibt man auch uv .

Beobachtung

$\cdot : \Sigma^* \times \Sigma^* \rightarrow \Sigma^*$ ist eine Abbildung (oder eine binäre Operation auf Σ^*) mit

- $(u \cdot v) \cdot w = u \cdot (v \cdot w)$ („Assoziativität“ - Klammern unnötig)
- $\varepsilon \cdot u = u \cdot \varepsilon = u$ („neutrales Element“- in Produkten unnötig).

Kürzer: $(\Sigma^*, \cdot, \varepsilon)$ ist ein **Monoid**.

$(\mathbb{N}, \cdot, 1)$ ist auch ein Monoid.

Definition

Für $w \in \Sigma^*$ und $n \in \mathbb{N}$ ist w^n induktiv definiert:

$$w^n = \begin{cases} \varepsilon & \text{falls } n = 0 \\ w \cdot w^{n-1} & \text{falls } n > 0 \end{cases}$$

Beispiel

$$\begin{aligned} 001100110011 &= 0011 \cdot 0011 \cdot 0011 = (0^2 1^2)^3 \\ abacabacabacabac &= abac \cdot abac \cdot abac \cdot abac \\ &= (abac)^4 \\ abacccc &= aba \cdot cccc \\ &= abac^4 \end{aligned}$$

d.h. die Potenzierung bindet stärker als die Verkettung.

Definition

Seien y, w Wörter über Σ . Dann heißt y

- **Präfix/Anfangsstück** von w , wenn es $z \in \Sigma^*$ gibt mit $yz = w$
- **Infix/Faktor** von w , wenn es $x, z \in \Sigma^*$ gibt mit $xyz = w$
- **Suffix/Endstück** von w , wenn es $x \in \Sigma^*$ gibt mit $xy = w$.

Beispiel

- aa ist Präfix von $aabba$ (und daher auch Faktor), aber kein Suffix
- aa ist kein Präfix, aber Suffix und Faktor von $abbaa$
- bb ist Faktor, aber weder Präfix noch Suffix von $aabba$
- aa ist Präfix, Faktor und Suffix von $aabbaa$
- Für jedes Wort w sind ε und w Präfix, Faktor und Suffix von w

Definition

- Sei Σ ein Alphabet.
Teilmenge von Σ^* werden **formale Sprachen über Σ** genannt.
- Eine Menge L ist eine **formale Sprache**, wenn es ein Alphabet Σ gibt, so daß L formale Sprache über Σ ist (d.h. so daß $L \subseteq \Sigma^*$).

Beispiel

1. Die Menge P_{10} der Dezimaldarstellungen von Primzahlen ist eine formale Sprache: das Alphabet ist $\{0, 1, 2, \dots, 9\}$
2. Die Menge P_2 der Binärdarstellungen von Primzahlen ist eine formale Sprache: das Alphabet ist $\{0, 1\}$.
Wegen $10 \in P_2 \setminus P_{10}$ gilt $P_2 \neq P_{10}$.
3. Die Menge P der Primzahlen ist keine formale Sprache, denn was sollte das Alphabet sein?

- Die Menge der syntaktisch korrekten deutschen Sätze - wirklich?

Nun schon wieder

Den eratmenden Schritt

Mühsam Berg hinauf!

Auf denn, nicht träge denn!

Strebend und hoffend hinan!

(Aus: J.W. v. Goethe, „An Schwager Kronos“)

Es ist unklar, ob es sich hier um korrekte deutsche Sätze handelt. Die „Menge aller grammatisch korrekten Sätze der deutschen Sprache“ ist gar keine Menge, da nicht eindeutig feststeht, ob eine Zeichenreihe dazugehört oder nicht. Also ist „sie“ auch keine formale Sprache.

Verabredung

Ab jetzt sagen wir oft einfach **Sprache** für „formale Sprache“.

Definition

Sind L_1 und L_2 Sprachen, so heißt die Sprache

$$L_1 L_2 = \{w \mid \exists w_1 \in L_1, w_2 \in L_2 : w = w_1 w_2\} \quad (\text{auch } L_1 \cdot L_2)$$

die **Konkatenation** oder **Verkettung** von L_1 und L_2 .

Beispiel

- $\{0\}^* \{1\}^* = \{0^i 1^j \mid i, j \geq 0\}$
- $\{0\} \cup \{1\} \{0, 1\}^*$ ist die Menge der Binärzahlen

Beobachtungen

- Die Verkettung von Sprachen ist assoziativ: $(L_1 L_2) L_3 = L_1 (L_2 L_3)$
- $\{\varepsilon\}$ ist neutrales Element: $L \{\varepsilon\} = \{\varepsilon\} L = L$
- $\emptyset$ ist auslöschendes Element: $\emptyset L = L \emptyset = \emptyset$.

Definition

Sei L Sprache und $n \in \mathbb{N}$. Dann ist L^n induktiv definiert:

$$L^n = \begin{cases} \{\varepsilon\} & \text{falls } n = 0 \\ L L^{n-1} & \text{falls } n > 0 \end{cases}$$

Es gilt also $L^n := \{w_1 w_2 \cdots w_n \mid w_1, \dots, w_n \in L\}$.

Beachte: $\{w^n \mid w \in L\} \subseteq L^n$, i.a. keine Gleichheit!

Beispiel

für $L = \{01, 0101\}$:

$01\,0101 \in L^2$, aber $01\,0101 \notin \{0101\,0101, 01\,01\} = \{w^2 \mid w \in L\}$

Definition

Sei L eine Sprache. Dann ist $L^* = \bigcup_{n \geq 0} L^n$ der **Kleene-Abschluß** oder die **Kleene-Iteration** von L .

Weiter ist $L^+ = \bigcup_{n \geq 1} L^n$ die **positive Iteration** von L .

Selbststudium: $L^+ = L \cdot L^* = L^* \cdot L$

Beachte: $\{w^n \mid w \in L, n \geq 0\} \subseteq L^*$, i.a. keine Gleichheit!

Beispiel

für $L = \{0, 11\}$: $011 \in L^* \setminus \{w^n \mid w \in L, n \geq 0\}$

Beobachtung

Sei Σ Alphabet.

- Sind L_1 und L_2 Sprachen über Σ , so auch
 - die Verkettung $L_1 L_2$,
 - die Kleene-Iteration L_1^* ,
 - die positive Iteration L_1^+ ,
 - die Vereinigung $L_1 \cup L_2$,
 - die Differenz $L_1 \setminus L_2$
 - und der Schnitt $L_1 \cap L_2$.
- $\emptyset$, Σ und Σ^* sind Sprachen über Σ .

Prioritätsregeln für Operationen auf Sprachen:

- Potenz/Iteration binden stärker als Konkatenation
- Konkatenation stärker als Vereinigung/Durchschnitt/Differenz.

Sprechweise: „Klasse“ von Sprachen (nicht „Menge“).

Beispiel

Die „Klasse aller Sprachen“,
die „Klasse der unendlichen Sprachen“,
die „Klasse der Sprachen über einem einelementigen Alphabet“,
die „Klasse der regulären Sprachen“, usw.

(Diese Gesamtheiten sind keine Mengen im strengen Sinne.)

Hinweise

1. am kommenden Donnerstag finden zwei AFS-Vorlesungen statt (Ersatz für Reformationstag): zusätzlich 17 Uhr, LdV-Hs1
2. Beispielhafte Testfragen aus dieser Vorlesung (andere sind möglich):
 - Was ist ein Alphabet?
 - Was ist Σ^* für eine Menge Σ ?
 - Was ist eine formale Sprache?
 - Was ist der Kleene-Abschluß einer Sprache L ?

Zusammenfassung 1. Vorlesung

in dieser Vorlesung neu

- Grundbegriffe: Alphabet, (formale) Sprache
- Operationen auf Sprachen

kommende Vorlesung

- grundlegende Methode zur Beschreibung formaler Sprachen
- Chomsky-Hierarchie