Formale Sprachen

Dozent: Prof. Dr. Michael Eichberg

Kontakt: michael.eichberg@dhbw.de, Raum 149B

Version: 1.0

Quelle: Die Folien sind teilweise inspiriert von oder basierend auf Lehrmaterial von Prof. Dr. Ritterbusch und Theoretische Informatik - kurzgefasst von Prof. Dr. Uwe Schöning.

Folien: https://delors.github.io/theo-algo-formale_sprachen/folien.de.rst.html

https://delors.github.io/theo-algo-formale_sprachen/folien.de.rst.html.pdf

Fehler melden: https://github.com/Delors/delors.github.io/issues

Einführung

Alphabete und Sprachen

Formale Sprachen sind ein zentraler Aspekt der theoretischen Informatik.

- Nutzungsinterface zwischen Computer und Mensch
- Grundlage für Programmiersprachen

Es gibt unterschiedliche Klassen und Modelle formaler Sprachen:

- Erkennbarkeit und Ausdruckskraft
- Anforderungen an Computermodelle zur Erkennbarkeit
- Komplexität von Verfahren zur Erkennung

Alphabete

Definition

Ein Alphabet $\Sigma = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ ist eine endliche Menge von Zeichen / Symbolen.

Beispiel

Abzählbare Mengen

- $\Sigma_{lat} = \{a, b, c, \dots, z\}$
- $\Sigma_{ziffer} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- $\Sigma_{unicode} = \{x | x \text{ ist ein Unicode-Zeichen}\}$
- $\Sigma_{logik} = \{0, 1, (,), \wedge, \vee, \neg, (,)\} \cup \Sigma_{lat}$

Kartesisches Produkt

Definition

Ein kartesisches Produkt wie $A \times B$ oder A^n für $n \in \mathbb{N}$ von Mengen oder Alphabeten bezeichnet die Menge der Tupel (a, b) oder $(a_1, \ldots, a_n)$ von Elementen der Mengen:

$$\begin{aligned} A \times B &:= & & \{(a,b) | a \in A, b \in B\} \\ A^n &:= \underbrace{A \times \ldots \times A}_{n \text{ Faktoren}} &=& \{(a1, \ldots, an) | a1, \ldots, an \in A\} \end{aligned}$$

Beispiel

- $\Sigma_{lat} \times \Sigma_{lat} = \{(a,a), (a,b), \ldots, (z,z)\}$
- $\Sigma_{lat}^3 = \{(a,a,a), (a,a,b), \ldots, (z,z,z)\}$

Kleene-Abschluss

Definition

Ein Wort ω ist ein endliches — ggf. leeres — Tupel $(w_1, w_2, \dots, w_n) \in \Sigma^n$ von Zeichen $w_k \in \Sigma$ eines Alphabets mit Länge $|\omega| = n$ der Anzahl der Zeichen.

- Wörter werden meist ohne Klammern geschrieben; d. h. $\omega = w_1 w_2 \dots w_n$.
- Das leere Wort (das Wort ohne Zeichen) wird mit ε bezeichnet.
- Besondere Wortmengen:
- $\Sigma^0 = \{\varepsilon\}$
- $\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n$
- $\Sigma^+ = \bigcup_{n=1}^{\infty} \Sigma^n$

Die Operationen M^* und M^+ auf einer Menge M werden als

- Kleene-$*$-Abschluss oder
- Kleene-$+$-Abschluss bezeichnet.

Beispiel

- $\Sigma_{lat}^* = \{\varepsilon, a, b, \dots, z, aa, ab, \dots, zz, aaa, \dots\}$
- $\Sigma_{lat}^+ = \{a, b, \dots, z, aa, ab, \dots, zz, aaa, \dots\}$

Beispiel

Sei $M = \{01, 2\}$, so ergeben sich u.a. diese Wortmengen:

$$\begin{aligned}
M^0 &= \varepsilon \\
M^1 &= 01, 2 \\
M^2 &= 0101, 012, 201, 22 \\
M^3 &= 010101, 01012, 01201, 0122, 20101, 2012, 2201, 222 \\
&\dots \\
M^+ &= M^1 \cup M^2 \cup \dots = 01, 2, 0101, 012, 201, 22, 010101, 01012, \dots \\
M^* &= M^0 \cup M^+ = \varepsilon, 01, 2, 0101, 012, 201, 22, 010101, 01012, \dots
\end{aligned}$$

Beobachtung

Die Wortlänge $|\omega|$ für ein $\omega \in L^*$ hängt von der Definition des Alphabets ab. So ist in diesem Beispiel $|222| = 3$ während $|0101| = 2$ ist.

Produkt und Konkatenation

Definition

Die Konkatenation von zwei Wörtern $\omega = (\omega_1, \ldots, \omega_n)$ und $\upsilon = (\upsilon_1, \ldots, \upsilon_m)$ ist definiert als das Wort, das durch ein aneinanderreihen der beiden Wörter entsteht:

$$\omega \cdot \upsilon = \omega\upsilon = (\omega_1, \ldots, \omega_n) \cdot (\upsilon_1, \ldots, \upsilon_m) = w_1 \ldots w_n \upsilon_1 \ldots \upsilon_m$$

Das leere Wort ist $\omega^0 = \varepsilon$ und die n-te Potenz von ω ist:

$$\omega^n = \underbrace{\omega \cdot \ldots \cdot \omega}_{n \text{ Faktoren}} \text{ für } n > 0$$

Beispiel

Sei $\Sigma = a, e, n, r$, sowie $\omega = \mathrm{na} \in \Sigma^*$ und $\upsilon = \mathrm{er} \in \Sigma^*$.

$\omega^2 = \mathrm{nana}$, $\upsilon\omega = \mathrm{erna}$ und $\upsilon\omega^2\upsilon = \mathrm{ernanaer}$

Abschluss-Eigenschaften

Bemerkung

Der Begriff *Abschluss in obiger Definition* bedeutet:

Auf einer Menge mit einer Verknüpfung liefert jede Anwendung der Operation mit Elementen wieder ein Element aus der Menge.

Beispiel

- die Subtraktion ist auf den natürlichen Zahlen nicht abgeschlossen,
- der Abschluss der natürlichen Zahlen bezüglich der Subtraktion sind die ganzen Zahlen.

Die Kleene-Abschlüsse und Multiplikationen werden später in regulären Ausdrücken auf Wörtern verwendet, damit ist dann der Abschluss oder das kartesische Produkt der Menge mit genau diesem Wort gemeint.

Beispiel

$$\begin{aligned} (ab)^+ &= \{ab\}^+ &&= \{ab, abab, ababab, \dots\} \\ cd^*e &= \{c\} \times \{d\}^* \times \{e\} &&= \{ce, cde, cdde, cddde, \dots\} \end{aligned}$$

Übung

Alphabet Σ = {a,el,en,g,l,ste}

Gegeben sei das Alphabet $\Sigma = \{a, el, en, g, l, ste\}$. Welche der folgenden Worte liegen in Σ^4?

ω_1 = galgen, ω_2 = stelle, ω_3 = sagen, ω_4 = lagen, ω_5 = allen, ω_6 = aalen

Alphabet Σ = {e,en,in,r,t,u}

Gegeben sei das Alphabet $\Sigma = e, en, in, r, t, u$. Welche der folgenden Worte liegen in Σ^5?

ω_1 = reiner, ω_2 = teurer, ω_3 = treuer, ω_4 = teuren, ω_5 = retten, ω_6 = teuer

Übung

Alphabet Σ = {e,g,in,l,s,ter}

Gegeben sei das Alphabet $\Sigma = e, g, in, l, s, ter$. Welche der folgenden Worte liegen in Σ^*?

ω_1 = tester, ω_2 = seile, ω_3 = lines, ω_4 = segel, ω_5 = seinen, ω_6 = erster

Formale Sprachen

Definition

Jede Teilmenge $L \subseteq \Sigma^*$ ist eine formale Sprache über dem Alphabet Σ.

Beispiel

Sei $\Sigma = \{0, 1, 2\}$, dann ist Σ^* die Menge oder Sprache von Wörtern aus den Ziffern 0, 1 oder 2 beliebiger Länge wie 101 oder auch 0001.

1

Die Menge $M \subset \Sigma^*$ der binären Zahlen ohne führende Nullen:

$$M = \{0\} \cup \{1\} \times \{0,1\}^* = \{0, 1, 10, 11, 100, 101, 110, 111, 1000, \ldots\}$$

2

Die Menge $M \subset \Sigma^*$ von einer gleichen Anzahl von 0 und 1 in dieser Reihenfolge:

$$M = \{0^n 1^n | n \in \mathbb{N}\} = \{01, 0011, 000111, 00001111, 0000011111, \ldots\}$$

3

Die Wörter $M \subset \Sigma^*$ mit gleicher Anzahl von 0, 1 und 2 in dieser Reihenfolge:

$$M = \{0^n 1^n 2^n | n \in \mathbb{N}\} = \{012, 001122, 000111222, 000011112222, \ldots\}$$

4

Die Menge $M \subset \Sigma^*$ mit Wörtern der Länge von Zweierpotenzen:

$$M = \{w \in \Sigma^* || w| = 2^n, n \in \mathbb{N}\} = \{0, 1, 2, 00, 01, \ldots, 21, 22, 0000, \ldots\}$$

Übung

Wörter bestimmen

Bestimmen Sie die Wörter der folgenden Sprache:

$L = \{acx^m(zq)^n | n \in \{0, 1\}, m \in \{1, 2\}\}$

Wörter bestimmen

Bestimmen Sie die Wörter der folgenden Sprache:

$L = \{(b^m a)^l za | m \in \{0, 1\}, l \in \{1, 2, 3\}\}$

Abzählbar und überabzählbar unendlich

Beobachtung

Selbst mit endlichen Alphabeten können formale Sprachen unendlich groß sein.

Definition

Eine Menge M ist *abzählbar*, wenn die einzelnen Elemente abzählbar sind, es also eine bijektive Funktion $f: N \to M$ von den natürlichen Zahlen $N = \mathbb{N}$ oder einer Teilmenge der natürlichen Zahlen $N \subset \mathbb{N}$ auf M gibt.

Wenn es keine solche Funktion geben kann, so ist die Menge *überabzählbar unendlich*.

Satz

Jede endliche Menge ist abzählbar.

Beweis

Eine endliche Menge M hat eine endliche Anzahl $n = |M|$ von Elementen.

Wird nun beginnend von $M_0 = M$ und $k = 1$ in n Schritten jeweils ein Element m_k der Menge M_{k-1} entnommen mit $M_k = M_{k-1}\{m_k\}$, so ist induktiv $|M_k| = |M_{k-1}| - 1 = n - k$ und es ist $M_n = \emptyset$.

Die Bijektion lautet dann $f: N \to M$ mit $f(k) = m_k$ mit $N = \{1, \dots, n\}$.

Satz

Jede Teilmenge $M \subseteq N$ einer abzählbaren Menge $N = \{n_1, n_2, \dots\}$ ist abzählbar.

Beweis

Sei $f(k) = n_k$ die Abzählung der Menge N. Sei $R = \{k \in \mathbb{N} | n_k \in M\}$; d. h. die Menge der Indizes der Elemente aus N, die in M sind. Dann ist die Einschränkung $f_{|R}: R \to M$ von f genau die Abzählung, die die Abzählbarkeit von M beweist.

Beispiel

Eine abzählbar unendliche Menge sind — zum Beispiel:

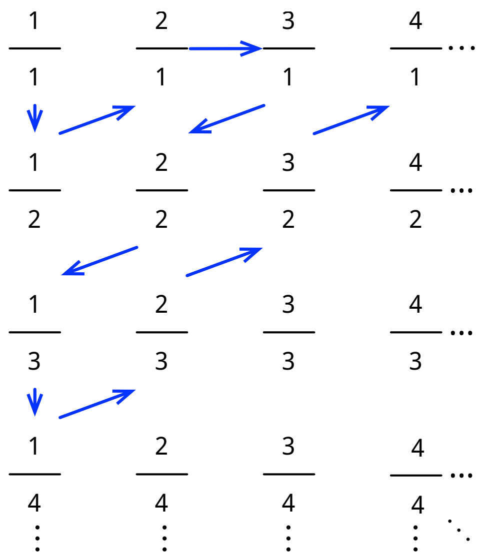
- die geraden Zahlen $\{2n | n \in \mathbb{N}\}$
- die Quadratzahlen $\{n^2 | n \in \mathbb{N}\}$
- die Menge der Fakultäten $\{n! | n \in \mathbb{N}\}$
- die ganzen Zahlen $\mathbb{Z}$ mit der Funktion:

$$f(n) = \begin{cases} n/2 & \text{für } n \text{ gerade} \\ -(n+1)/2 & \text{für } n \text{ ungerade} \end{cases}$$

$$f(1) = 0,\ f(2) = 1,\ f(3) = -1,\ f(4) = 2,\ f(5) = -2,\ \dots$$

Beispiel

Die rationalen Zahlen $\mathbb{Q}$ sind abzählbar unendlich.



Rationale Zahlen können als Brüche dargestellt werden und mit Hilfe des Diagonalisierungsverfahren von Cantor in eine Bijektion zu den natürlichen Zahlen gebracht werden.

Die 0 und alle negativen Brüche können wie zuvor eingeschoben werden. Auch alle rationalen Vektoren $\mathbb{Q}^n$ in beliebiger Dimension $n \in \mathbb{N}$ sind so abzählbar.

Satz

Für jede endliche Menge oder Alphabet Σ ist deren Kleene-Abschluss Σ^* abzählbar.

Beweis

Ist das Alphabet Σ leer, so ist auch Σ^* leer, und damit für $N = \emptyset$ trivial abzählbar.

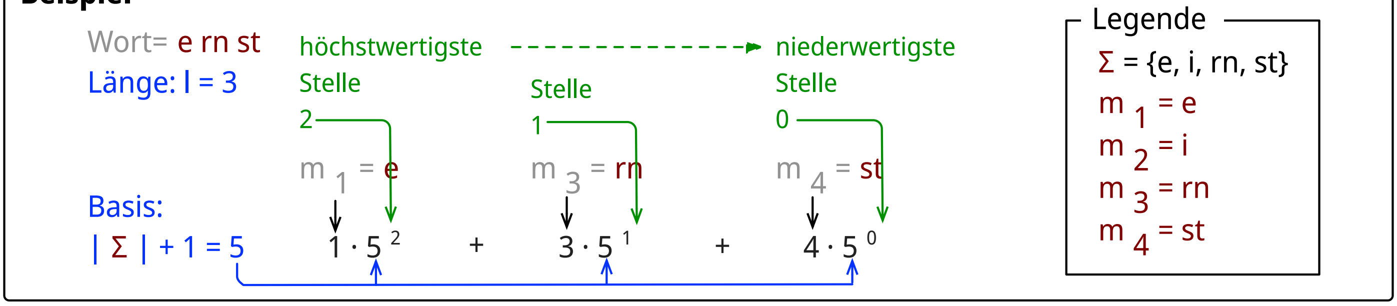
Ist Σ nicht leer, dann besitzt Σ mit Größe $n = |\Sigma|$ eine Aufzählung m_k mit $k = 1, \ldots, n$.

Jedes Wort $w = m_{k_1} m_{k_2} \ldots m_{k_l}$ kann dann im Stellenwertsystem zur Basis $n + 1$ dargestellt werden:

$$1 + k_1 \cdot (n+1)^{l-1} + k_2(n+1)^{l-2} + \ldots + k_l(n+1)^0$$

und somit der Zahl $1 + (k_1 k_2 \ldots k_l)_{(n+1)}$[1] zugeordnet werden.

Beispiel



Die Abbildung $f : N \to \Sigma^*$ mit $N \subseteq \mathbb{N}$ ergibt sich für $f(x)$ aus der Stellenwertdarstellung von $x - 1 > 0$ zur Basis $n + 1$ beginnend mit der höchstwertigen Ziffer k_1 bis zur letzten Stelle k_l.

Das Bild $f(x)$ ist dann das Wort $m_{k_1} m_{k_2} \ldots m_{k_l}$.

Das leere Wort ε wird von 1 abgebildet und entsprechend ist $f(1) = \varepsilon$.

Beispiel

Sei $\Sigma = \{e, i, rn, st\}$ mit Aufzählung $m_1 = e$, $m_2 = i$, $m_3 = rn$, $m_4 = st$, dann haben die folgenden Wörter diese Abzählung nach Stellenwert:

x	1	2	3	4	5
	1				

		$1 \quad + \quad 1$ $1_5 \quad + \quad 1_5$ [2]	$1 \quad + \quad 2$ $1_5 \quad + \quad 2_5$	$1 \quad + \quad 3$ $1_5 \quad + \quad 3_5$	$1 \quad + \quad 4$ $1_5 \quad + \quad 4_5$
Wort	ϵ	e	i	rn	st
$f(x)$	$f(1) = \epsilon$	$f(2) = e$	$f(3) = i$	$f(4) = rn$	$f(5) = st$ *(Anm.: k ist 4 für st)*

x	...	$7 = 1 + 6$ $12_5 = 1_5 + 11_5$	$8 = 1 + 7$ $13_5 = 1_5 + 12_5$	...	$45 = 1 + 44$ $140_5 = 1_5 + 134_5$	...
	...	$1 \quad + \quad 1 \cdot 5 + 1$ $1 \quad + \quad 11_5$	$1 \quad + \quad 1 \cdot 5 + 2$ $1 \quad + \quad 12_5$	...	$1 \quad + \quad 1 \cdot 25 + 3 \cdot 5 + 4$ $1 \quad + \quad 134_5$	...
Wort	...	ee	ei	...	ernst	...

Unbesetzt bleibt, wo eine 0 in der Stellenwertdarstellung vorliegt. Zum Beispiel ist $f(6) = 1 + 1 \cdot 5^1 + 0 \cdot 5^0 = 1_5 + 10_5$.

Satz

Jede formale Sprache is abzählbar.

Beweis

Da jede formale Sprache L über einem endlichen Alphabet Σ definiert ist, ist das eine direkte Folge aus vorherigem Satz, dass Σ^* abzählbar ist, und wie zuvor gezeigt damit auch die Teilmenge $L \subseteq \Sigma^*$ abzählbar ist.

Abzählen mit Hilfe von Gödelnummern

Definition

Sei (p_n) die Folge der Primzahlen:

$p_1 = 2, p_2 = 3, p_3 = 5, p_4 = 7, p_5 = 11, p_6 = 13, \ldots$

Für eine abzählbare Menge $M = m_1, m_2, \ldots$ ist die Gödelnummer $c_M : M^* \to \mathbb{N}$ des Tupels $w = (m_{k_1}, m_{k_2}, \ldots, m_{k_l})$ gegeben durch

$$c_M(w) = p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_l^{k_l} = \prod_{i=1}^{l} p_i^{k_i}$$

Beispiel

Sei $\Sigma = \{e, i, rn, st\}$ mit Aufzählung $m_1 = e$, $m_2 = i$, $m_3 = rn$, $m_4 = st$, dann haben die folgenden Wörter diese Gödelnummern:

w	e	i	rn	st	$ernst$
$c_M(w)$	2^1	3^1	5^2	7^3	$2^1 \cdot 3^1 \cdot 5^2 \cdot 7^3 = 2 \cdot 3 \cdot 25 \cdot 343 = 51450$

Satz

Die Menge von endlichen Folgen $P = \{p = (w_1, w_2, \ldots, w_n) | w_k \in L, n \in \mathbb{N}\}$ aus Wörtern einer formalen Sprache $L \subseteq \Sigma^*$ (also Programmen) über einem Alphabet Σ ist abzählbar.

Beweis

Jede formale Sprache $L \subseteq \Sigma^*$ ist abzählbar. Damit kann nach Definition für jede Folge $p \in P$ injektiv eine Gödelnummer $c_L(p)$ über L bestimmt werden. Auf der Menge $N = \{x = c_L(p) | p \in P\}$ kann die

Umkehrung $f : N \to P$ von c_L auf P eingeschränkten bijektiven Funktion $c_{L|P} : P \to N$ bestimmt werden, und damit ist P abzählbar.

---

[2] Wir haben immer $1 + \ldots$, da wir noch das leere Wort ε haben.

[1] Die Darstellung $(k_1 k_2 \ldots k_l)_{(n+1)}$ ist die Stellenwertdarstellung zur Basis $n+1$ des Wortes w.