Theoretische Informatik

Alphabete, Wörter und Sprachen

Alphabet

Ein Alphabet ist eine endliche, nichtleere Menge von Symbolen. Bsp.:

$$\Sigma = \{a, b, c\}, \quad \Sigma_{\text{bool}} = \{0, 1\}, \quad \Sigma_{\text{lat}} = \{a, b, c \cdots, z\}$$

Wörter

Ein Wort über einem Alphabet ist eine endliche Folge von Symbolen aus diesem Alphabet. Das leere Wort wird mit ε bezeichnet.

$$w = (x_1, x_2, \ldots, x_n), \quad x_i \in \Sigma, \quad n \in \mathbb{N}$$

Die Länge eines Wortes w ist die Anzahl der Symbole, also $|w| = n$. Die absolute Häufigkeit eines Symbols a in einem Wort w wird mit $|w|_a$ bezeichnet. Zusätzlich gilt:

$$|\varepsilon| = |\varepsilon|_a = 0, \quad a \in \Sigma$$

Spiegelung und Palindrome

Mit w^R wird das Spiegelwort von w bezeichnet, also die Umkehrung der Symbolfolge.

$$w^R = (x_1, x_2, \ldots, x_n)^R = (x_n, x_{n-1}, \ldots, x_1)$$

Wenn $w = w^R$ gilt, dann ist w ein Palindrom.

Wörter der Länge k

Die Menge aller Wörter der Länge k über einem Alphabet Σ wird mit Σ^k bezeichnet:

$$\Sigma^k = \{w \in \Sigma^* \mid |w| = k\}$$

Unabhängig von Σ gilt stets $\Sigma^0 = \{\varepsilon\}$.

Kleenesche Hülle

Die Menge aller Wörter über einem Alphabet Σ wird mit Σ^* bezeichnet:

$$\Sigma^* = \bigcup_{k \geq 0} \Sigma^k$$

Die Menge aller nichtleeren Wörter über Σ wird mit Σ^+ bezeichnet:

$$\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$$

Teilwortrelationen

Infix v ist Infix von w $\iff$ Es existieren $x, y \in \Sigma^*$ mit $w = xvy$.

Präfix v ist Präfix von w $\iff$ Es existiert $y \in \Sigma^*$ mit $w = vy$.

Suffix v ist Suffix von w $\iff$ Es existiert $x \in \Sigma^*$ mit $w = xv$.

Man spricht von einem *echten* Infix/Präfix/Suffix, wenn $v \neq w$ gilt.

Konkatination

Die Konkatenation von zwei Wörtern x und y ist die Aneinanderreihung der Symbole von x gefolgt von den Symbolen von y:

$$x \circ y = xy = (x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m)$$

Die Länge der Konkatenation ist die Summe der Längen der beiden Wörter:

$$|xy| = |x| + |y|$$

Die Konkatenation mit dem leeren Wort ε hat keine Auswirkung auf das Wort:

$$w\varepsilon = \varepsilon w = w$$

Wortpotenzen

Die n-te Potenz eines Wortes x wird für alle $n \in \mathbb{N}$ definiert als

$$x^0 := \varepsilon$$
$$x^{n+1} := x^n x$$

Die Potenzierung mit der Kleeneschen Hülle ergibt die gleiche Menge:

$$(A^*)^* = A^*$$

Sprache

Eine Teilmenge $L \subseteq \Sigma^*$ von Wörtern heisst Sprache über dem Alphabet Σ. Es gelten die folgenden Eigenschaften:

- $\{\} = \emptyset$ ist die leere Sprache für jedes Alphabet Σ.
- $\{\varepsilon\}$ ist die Sprache, die nur das leere Wort enthält, für jedes Alphabet Σ.
- Σ^* ist die Sprache aller Wörter über Σ.

Konkatenation

Die Konkatenation von Sprachen A und B ist definiert als die Menge aller Konkatenationen von Wörtern aus A mit Wörtern aus B:

$$AB = \{uv \mid u \in A, v \in B\}$$

Die Konkatenation von Sprachen ist assoziativ, aber im Allgemeinen nicht kommutativ:

$$(AB)C = A(BC), \quad AB \neq BA$$

Ist A eine Sprache über Σ und B eine Sprache über Γ, so ist AB eine Sprache über $\Sigma \cup \Gamma$.

Kleenesche Hülle

Die Kleenesche Hülle A^* einer Sprache A ist definiert als die Vereinigung aller Potenzen von A:

$$A^* = \bigcup_{k \geq 0} A^k$$

Dabei ist $A^0 = \{\varepsilon\}$ und $A^{k+1} = A^k A$. Die Kleenesche Hülle erfüllt die Eigenschaft:

$$(A^*)^* = A^*$$

Entscheidungsproblem

Sei eine Sprache $L \subseteq \Sigma^*$ gegeben. Das Entscheidungsproblem für L besteht darin, für ein beliebiges Wort $x \in \Sigma^*$ zu entscheiden, ob x in L enthalten ist oder nicht:

$$x \in \Sigma^* \mapsto \begin{cases} \text{JA} & \text{wenn } x \in L, \\ \text{NEIN} & \text{wenn } x \notin L. \end{cases}$$

By Justin Iven Müller (2026)
github.com/JustinIven/zhaw-cheatsheets