

1 Wörter, Alphabete und Sprachen

1.1 Alphabet

Ein Alphabet ist definiert als eine endliche, nicht leere Menge von Symbolen. Alphabete werden in der Regel durch griechische Grossbuchstaben (meist Σ oder Γ) dargestellt.

1.2 Wort

Ein Wort ist eine endliche Symbolabfolge bestehend aus Symbolen eines Alphabets. Wörter werden mit einem kleinen Buchstaben bezeichnet. Das leere Wort enthält keine Symbole und wird durch ε dargestellt. Die Länge eines Wortes w beschreibt die Anzahl Symbole die dieses enthält und wird mit Betragstrichen visualisiert. Die Häufigkeit eines bestimmten Symbols x in einem Wort w kann mithilfe von $|w|_x$ dargestellt werden. Ein hochgestelltes R eines Wortes entspricht dem Rückwärts gelesenen Wort. Wenn $w = w^R \Rightarrow w$ ist ein Palindrom.

1.2.1 Teilwörter

Ein Teilwort kann entweder am Wortanfang (Präfix), in der Wortmitte (Infix) oder am Wortende (Suffix) stehen. Es wird unterschieden zwischen *echten* und *nicht echten* Teilwörtern unterschieden, wobei ein echtes Teilwort nicht dem ganzen Wort entspricht. Zu beachten ist, dass ε ein Teilwort jedes Wortes ist, sowohl als Präfix, Infix und auch als Suffix.

Die Menge aller Wörter mit der Länge k über das Alphabet Σ wird als Σ^k beschrieben. $\Sigma^0 = \{\varepsilon\}$. Alle Wörter eines Alphabets Σ wird als Σ^* bezeichnet. Σ^+ entspricht $\Sigma^* \setminus \{\varepsilon\}$.

Wörter können aneinander gehängt werden. Diesen Vorgang nennt man **Konkatenation**. Dabei gilt $w \circ v = wv$. Wortpotenzen konkatenieren das gleiche Wort n mal aneinander. $a^3 = aaa$

1.3 Sprache

Eine Sprache definiert eine Teilmenge aller möglichen Wörter eines Alphabets. Dabei können Sprachen unendliche viele Wörter enthalten, wobei die Wörter endliche Symbolabfolgen eines endlichen Alphabets sind. Sprachen können gleich den Wörtern konkateniert werden und somit eine neue Sprache bilden.

2 Regex

Regex (= regular expression) ist eine Ausdrucksweise um valide Wörter einer Sprache zu beschreiben. Mit regulären Ausdrücken kann man Muster beschreiben, welche programmatisch auf Strings angewendet werden können.

Tabelle 1: Regex-Operatoren

Operator	Funktion
()	Gruppert mehrere Symbole um weitere Operationen auf die Gesamte Zeichenkette anzuwenden
	Entspricht einem "oder"
*	Bedeutet "beliebig viele" des gefolgten Elements
+	Bedeutet "mindestens einmal" das gefolgte Element
[]	Gruppert mehrere Elemente, von denen eines zutreffen muss
^	Markiert den Anfang eines Strings
\$	Markiert das Ende eines Strings
?	Bedeutet "Ein oder Null mal" das gefolgte Element