

„Einführung in den Umgang mit regulären Ausdrücken“

Sitzung 2

Thorsten Fritze

t.fritze@ub.uni-frankfurt.de

UB Frankfurt / Fachinformationsdienst Linguistik

8. Februar 2024

Gliederung

- 1 Wiederholung

Was sind Reguläre Ausdrücke

Zweck & Verwendung

Reguläre Ausdrücke erlauben es, Zeichenkettenmuster zu beschreiben und diese zum Suchen/Finden, Ersetzen, Validieren, etc. zu verwenden.

Metazeichen	Bedeutung
<code>^</code>	Zeichettenanfang
<code>\$</code>	Zeichenkettenende
<code>.</code>	Beliebiges Zeichen
<code>[a-zA-Z0-9]</code>	Zeichenauswahl
<code>[^a-zA-Z0-9]</code>	Negierte Zeichenauswahl
<code>\s</code>	Leerzeichen & Tabulatoren
<code>\d</code>	Ziffern 0-9
<code>\w</code>	Buchstaben, Ziffern, ggf. Bindestrich
<code>\S</code>	Alles außer Leerzeichen & Tabulatoren
<code>\D</code>	Alles außer Ziffern 0-9
<code>\W</code>	Alles außer Buchstaben, Ziffern, ggf. Bindestrich
<code>*</code>	Null- oder mehrmals
<code>?</code>	Null- oder einmal
<code>+</code>	Ein- oder mehrmals
<code>{5}</code>	genau fünf Mal
<code>{2, 5}</code>	Zwei bis fünf Mal
<code>{5, }</code>	Fünf Mal oder mehr

Tabelle: Bestandteile regulärer Ausdrücke (in Auswahl)

Software

Eine kleine Auswahl an Programmen, die Unterstützung für RegEx anbieten:

- **Office-Pakete:** LibreOffice
- **Textverarbeitung:** MS Word (eingeschränkt + abweichende Syntax)
- **Texteditoren:** Notepad++, Textmate, Geany ...
- **Programmiersprachen:** R, Python (und alle anderen Programmiersprachen)
- **Datenanalyse/-konvertierung:** AntConc, OpenRefine, ...
- **Standardwerkzeuge unter Linux/Unix:** sed, awk, grep, ...