Regular Expressions

Informatik
4. Jahr, 2. Semester
Kantonsschule am Burggraben

Ivo Blöchliger

Regular Expressions

- Beschreibung von Text-Mustern
- Beispiel Datum in der Form 12-2-2021
- 1-2 Ziffern, 1 Minus, 1-2 Ziffern, 1 Minus, 4 Ziffern
- \d{1,2}-\d{1,2}-\d{4}
  - \d ist eine Abkürzung für [0-9]
  - {n,m} heisst mindestens n, höchstens m Mal das zuvor.
  - Die meisten Zeichen stehen für sich selbst, wie z.B. - (Minus)

Anwendung Regular Expressions

- Passt ein Muster (kann ein Datum sein)?
- Muster suchen und extrahieren (alle Daten in einem Text)
- Muster suchen und ersetzen

- Datenextraktion und Aufbereitung, damit diese in ein weiteres Programm importiert werden können (z.B. Tabellenkalkulation).

Wichtigste Sonderzeichen

- . (Punkt): Steht für genau ein beliebiges Zeichen
- \ (Backslash): Das nächste Zeichen speziell behandeln
  - \d (Ziffer), \. (Punkt), \\ (Backslash), \n (Zeilenumbruch)
- [0-9a-f] (Auswahl): Genau eines der Zeichen
- [^0-9]: Alles ausser den Zeichen
- | (Pipe): Auswahl, z.B. Mo|Di|Mi|Do
  - in diesem speziellen Fall das gleiche wie [MD][io]

Quantifier

- *: Beliebig viele (maximal) des vorigen Ausdrucks
- *?: Beliebig viele (minimal)
- ?: Null oder einmal
- +: Ein oder mehr mal (maximal)
- {5}: Genau 5 mal
- {3,8}: zwischen 3 und 8 Mal

Viele Beispiele eieiei!

- **.ie.**
  ”Viel”, “piel”, “eiei”
- **e.*e**
  ”ele Beispiele eieie”
- **e.*?e**
  ”ele”, “ele”, “eie”
- [A-Z][a-z]*
  ”Viele”, “Beispiele”

Gruppen

- Runde Klammern öffnen und schliessen Gruppen
- Werden als separate Matches gespeichert, nummeriert in der Reihenfolge öffnender Klammern.
- Praktisch für die Extraktion oder für die Substitution
- Werden in $1, $2 etc. gespeichert
  - Je nach Programm manchmal auch \1, \2, etc.

Link und Text extrahieren

- Hello
- (*.?)
- Ersetzen durch z.B. "$2";"$1" für Text und Link

Beliebige Zeichen, so wenig wie möchlich, gefolgt von "

Es könnte noch etwas stehen hier, aber kein >

Die Matches in den runden Klammern werden in $1, $2 gespeichert und können in einer Ersetzung benutzt werden.