



# Reguläre Ausdrücke

Robert Schulze (rob@rob-schulze.de)

1.2.2007



# Begriffserklärung

- abstrakte Beschreibung von Zeichenketten in beliebig strengem Umfang
- kann u.a. enthalten:
  - lexikalische Zeichen (Buchstaben, Zahlen)
  - den Punkt (.)
  - Zeichenbereiche ([0-9])
  - Quantoren (\*, +, ?, {n}, {min,}, {,max}, {min,max})
  - Anker (\$, ^)



# Aufgabenfelder

- finden/extrahieren von Zeichenketten
- suchen & ersetzen in Zeichenketten
- Überprüfung auf Gültigkeit für (alle) textbasierten Sprachen



# Syntax: einfaches Suchen

- rob 'rob' irgendwo
- ^rob 'rob' am Anfang
- rob\$ 'rob' am Ende
- ^rob\$ nur 'rob'
- \^rob\\$ '^rob\$' irgendwo
- .ob\$ irgendwas und 'ob' am Ende
- rob|p rob oder rop



# Syntax: Auswahlen (Klassen)

- [01] 0 oder 1
- [0-9] beliebige Ziffer
- [A-Za-z0-9] alle Buchstaben/Ziffern
- [^A] kein großes A
- [\. \-] Punkt oder Minus



# Syntax: Quantoren

- $a^*$       kein oder beliebige viele 'a'
- $a?$       kein oder ein a
- $ab^+c$       a, beliebig viele b, c
- $a\{1\}$       genau 1 a
- $ab\{2,\}c$       a, mindestens 2 b, c
- $abc\{2,5\}$       a, b, (cc, ccc, cccc, ccccc)
- $ab\{,5\}c$       a, höchstens 5 b, c



# Zusammen sind wir stark

- hexadezimaler Wert ( $\leq 255$ )
  - `[A-Za-z0-9]{1,2}`
- Beliebiger Satz, der mit einem großen Z beginnt
  - `Z[^\.\?\\!]+(\.\|\\?|\\!)`
- Tier, Hier oder Vier?
  - `[THV]ier`



# Zusammen sind wir stark

- Währungsangabe
  - $[0-9,]^+$  (DM|EUR)
- Datum
  - $[0-9]^+\backslash.[0-9]^+\backslash.[0-9]\{4\}$
- eine beliebige Menge vielerlei Nahrung
  - $[0-9]^+ k?g$  (Bock|Pferde|)wurst(|salat)





# Regex in Perl

- Suchen: `$var=~m/REGEX/FLAGS`
- Ersetzen: `$var=~s/REGEX/SUBST/FLAGS`
- Flags:
  - `g` globale Suche so oft wie möglich
  - `s` Argument als einzeilig behandeln
  - `c` Suche nie zurücksetzen
  - `i` “case-insensitive” suchen



# Regex in PHP

- Suchen:
  - `preg_match(REGEX,VAR)`
  - `preg_match_all(REGEX,VAR)`
- Ersetzen:
  - `preg_replace(REGEX,SUBST,VAR)`



# Regex in Javascript

- Suchen:
  - `STRING.search(/AUSDRUCK/)`
- Ersetzen:
  - `STRING.replace(/AUSDRUCK/,STRING)`
- Flags wie in Perl möglich!



# Quellen

<http://www.wikipedia.de/Regex>

für PHP: <http://www.php.net/>

für Javascript: <http://de.selfhtml.org/>



# Dank

s/^/Vielen /

s/ / herzlichen /

s/\$/!