

Zusammenfassung Blatt 4

Regular Expression

Konstrukt	Beschreibung	Muster	Übereinstimmung
[<i>character_group</i>]	Entspricht einem beliebigen einzelnen Zeichen in <i>character_group</i> . Bei der Entsprechung wird standardmäßig die Groß- und Kleinschreibung berücksichtigt.	[ae]	"a" in "wage"
			"a", "e" in "klasse"
[^ <i>character_group</i>]	Negation: Entspricht jedem beliebigen einzelnen Zeichen, das nicht in <i>character_group</i> enthalten ist. Standardmäßig wird bei Zeichen in <i>character_group</i> die Groß-/Kleinschreibung beachtet.	[^aei]	"r", "g", "n" in "ringen"
[<i>first-last</i>]	Zeichenbereich: Entspricht jedem beliebigen einzelnen Zeichen im Bereich von <i>first</i> bis <i>last</i> .	[A-Z]	"A", "B" in "AB123"
.	Platzhalterzeichen: Entspricht jedem beliebigen einzelnen Zeichen außer \n.	a.e	"ade" in "gerade"
	Damit es einem Punkt als Literalzeichen entspricht, muss ihm ein Escapezeichen (\.) vorangestellt werden.		"ase" in "klasse"
\w	Entspricht einem beliebigen Wortzeichen.	\w	"I", "D", "A", "1", "3" in "ID A1.3"
\W	Entspricht einem beliebigen Nichtwortzeichen.	\W	" ", "." in "ID A1.3"
\d	Entspricht einer beliebigen Dezimalziffer.	\d	"4" in "4 = IV"
\D	Entspricht einem beliebigen Zeichen, das keine Dezimalziffer ist.	\D	" ", "=", " ", "I", "V" in "4 = IV"
*	Entspricht dem vorangehenden Element nicht oder mehrmals.	\d*\.\d	".0", "19.9", "219.9"

+	Entspricht dem vorangehenden Element einmal oder mehrmals.	"be+"	"bee" in "beeilt", "be" in "bei"
?	Entspricht dem vorangehenden Element nicht oder einmal.	"rai?n"	"ran", "rain"
{ <i>n</i> }	Entspricht dem vorangehenden Element genau <i>n</i> -mal.	"\d{3}"	",043" in "1,043.6", ",876", ",543" und ",210" in "9,876,543,210"
{ <i>n</i> ,}	Entspricht dem vorangehenden Element mindestens <i>n</i> -mal.	"\d{2,}"	"166", "29", "1930"
{ <i>n</i> , <i>m</i> }	Entspricht dem vorangehenden Element mindestens <i>n</i> -, höchstens jedoch <i>m</i> -mal.	"\d{3,5}"	"166", "17668"
			"19302" in "193024"
	Entspricht jedem beliebigen durch einen senkrechten Strich () getrennten Element.	th(e is at)	"the", "this" in "This is the day."

Backus Naur Form (BNF)

Beschreibt im Prinzip den Aufbau. Es wird immer weiter ersetzt bis man bei Terminalsymbolen angekommen ist (z.B. die Zahlen 0|1|2,...) die nicht ersetzt werden können. Am einfachsten versteht man dies mit Hilfe eines Beispiels.

Beispiel:

Personennamen wie Hr. Prof. Dr. Stefan Max Holger von Mustermann

<Personenname> ::= <anrede><space><namen><space><adelstitel><space><nachname>

<anrede> ::= <geschlechtsbezeichner> | <geschlechtsbezeichner> <space><titels>

<geschlechtsbezeichner> ::= "Hr." | "Fr."

<titels> ::= <titel> | <titel> <space> <titels>

<titel> ::= "Prof" | "Dr." | "Habil." | "Dipl."

<namen> ::= <name> | <name> <space> <namen>

<name> ::= "Stefan" | "Max" | "Holger" | "Tim" | "Uwe"

<space> ::= " "

<adelstitel> ::= "" | "von"

<nachname> ::= "Mustermann" | "Musterfrau" | "Müller" | "Merkel"

Erklärung:

In <Personenname> wird genau Definiert wie der Name aufgebaut ist. Zuerst die Anrede, dann die einzelnen (Vor)namen, dann ggfls. ein Titel und zuletzt der Nachname. Die einzelnen Symbole werden weiter definiert bis sie zu Terminalsymbolen werden (wie z.b. in titel: Prof).

Anmerkung zum Beispiel: Die Symbole "namen" und "titels" könnten ggfls. zu Missverständnissen führen (vorallem titelS). Hierbei handelt es sich nicht um Schreibfehler sondern um frei gewählte Bezeichner. Da es Menschen mit mehreren Vornamen/Titeln gibt muss dies genauer definiert werden. Hierfür wurden die "Oberbegriffe" <namen> und <titels> gewählt die später genauer definiert werden. (<namen> ::= <name> | <name><space><namen> → entweder ein Name einsetzen oder ein Name und beliebig viele weiteren <namen>)

Beispiele für akzeptierte Ausdrücke:

Hr. Stefan Müller, Fr. Holger Merkel, Fr. Prof. Dipl. Max Holger Tim Uwe Stefan Max Musterfrau