

---

# Reguliere expressies: Theorie I

Huub de Beer

Eindhoven, 4 juni 2011

---

## Patronen eenvoudig?

- ▶ Patronen zijn overal: je kent ze goed
- ▶ En daardoor lijken ze zo eenvoudig
- ▶ Maar als je de patronen gaat beschrijven, blijken ze niet zo eenvoudig:
  - ▶ precisie is belangrijk
  - ▶ in woorden worden patronen al snel lang
  - ▶ het is lastig om alle uitzonderingen en speciale gevallen ook op te nemen
- ▶ Oplossing: reguliere expressies

---

# Enkele basisdefinities: alfabet

## Alfabet

Een verzameling unieke symbolen noemen we een **alfabet**

## Voorbeelden

- ▶  $\mathbb{C} = \text{'0', '1', '2', '3', '4', '5', '6', '7', '8', '9'}$
- ▶  $\mathbb{G} = \text{'a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k', 'l', 'm', 'n', 'o', 'p', 'q', 'r', 's', 't', 'u', 'v', 'w', 'x', 'y', 'z', '0', '1', '2', '3', '4', '5', '6', '7', '8', '9'}$

---

# Enkele basisdefinities: alfabet

## Alfabet

Een verzameling unieke symbolen noemen we een **alfabet**

## Voorbeelden

- ▶  $\mathbb{C} = \text{'0', '1', '2', '3', '4', '5', '6', '7', '8', '9'}$
- ▶  $\mathbb{G} = \text{'a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k', 'l', 'm', 'n', 'o', 'p', 'q', 'r', 's', 't', 'u', 'v', 'w', 'x', 'y', 'z', '0', '1', '2', '3', '4', '5', '6', '7', '8', '9'}$

---

## Enkele basisdefinities: zin

### Zin

een reeks aaneengesloten symbolen uit een alfabet  $\mathbb{A}$   
noemen we een **zin** over alfabet  $\mathbb{A}$ .

### Voorbeelden

Gegeven alfabet  $\mathbb{G}$ .

- ▶ 'aap'
- ▶ '12konijnen'
- ▶ (fout) 'dit is lastig'
- ▶ (fout) '06-12552389'

---

## Enkele basisdefinities: zin

### Zin

een reeks aaneengesloten symbolen uit een alfabet  $\mathbb{A}$   
noemen we een **zin** over alfabet  $\mathbb{A}$ .

### Voorbeelden

Gegeven alfabet  $\mathbb{G}$ .

- ▶ 'aap'
- ▶ '12konijnen'
- ▶ (fout) 'dit is lastig'
- ▶ (fout) '06-12552389'

---

## Enkele basisdefinities: taal

### Taal

Een verzameling zinnen gemaakt met symbolen uit een alfabet  $\mathbb{A}$  noemen we een **taal** over alfabet  $\mathbb{A}$ . Vaak worden eisen gesteld waaraan zinnen in een taal moeten voldoen: niet alle combinaties van symbolen in  $\mathbb{A}$  zijn dan zinnen in de taal.

### Voorbeelden

Gegeven alfabet  $\mathbb{G}$ .

- ▶ ..., 'aaaap', 'aaap', 'aap', 'ap', 'p': nul of meer a's gevolgd door een p.
- ▶ 'jan': de zin 'jan'

---

## Enkele basisdefinities: taal

### Taal

Een verzameling zinnen gemaakt met symbolen uit een alfabet  $\mathbb{A}$  noemen we een **taal** over alfabet  $\mathbb{A}$ . Vaak worden eisen gesteld waaraan zinnen in een taal moeten voldoen: niet alle combinaties van symbolen in  $\mathbb{A}$  zijn dan zinnen in de taal.

### Voorbeelden

Gegeven alfabet  $\mathbb{G}$ .

- ▶ ..., 'aaaap', 'aaap', 'aap', 'ap', 'p': nul of meer a's gevolgd door een p.
- ▶ 'jan': de zin 'jan'



---

## Voorbeeld, zinnen beginnend met een 'a'

### Alfabet $\mathbb{A}$

alfabet  $\mathbb{A} = \text{'a'}, \text{'b' en 'c'}$

### Taal $\mathcal{L}()$

De taal  $\mathcal{L}()$  over alfabet  $\mathbb{A}$  bestaat uit alle zinnen (over  $\mathbb{A}$ ) die met een 'a' beginnen. 'aab', 'a', 'aaaaaaa', 'acba' en 'acccc' zijn dus zinnen in de taal  $\mathcal{L}()$ . 'a4', '23' en 'Aab' zijn geen zinnen in de taal  $\mathcal{L}()$

### Reguliere expressie

$a \cdot (a|b|c)^*$

---

## Voorbeeld, zinnen beginnend met een 'a'

### Alfabet $\mathbb{A}$

alfabet  $\mathbb{A} = \text{'a'}, \text{'b'} \text{ en } \text{'c'}$

### Taal $\mathcal{L}()$

De taal  $\mathcal{L}()$  over alfabet  $\mathbb{A}$  bestaat uit alle zinnen (over  $\mathbb{A}$ ) die met een 'a' beginnen. 'aab', 'a', 'aaaaaaa', 'acba' en 'acccc' zijn dus zinnen in de taal  $\mathcal{L}()$ . 'a4', '23' en 'Aab' zijn geen zinnen in de taal  $\mathcal{L}()$

### Reguliere expressie

$a \cdot (a|b|c)^*$

---

# Definitie reguliere expressie

## Reguliere expressie

Gegeven alfabet  $\mathbb{A}$ . De volgende expressies zijn reguliere expressies over alfabet  $\mathbb{A}$ :

- ▶ Elk karakterer uit  $\mathbb{A}$  is een reguliere expressie. We noemen dit soort reguliere expressies ook wel ***primaire reguliere expressies***.
- ▶ Als  $r$  en  $s$  reguliere expressies zijn, dan zijn
  - ▶  $r \cdot s$
  - ▶  $r|s$ ,
  - ▶  $r^*$  en
  - ▶  $(r)$

dat ook.

Elke reguliere expressie specificeert nul of meer zinnen over alfabet  $\mathbb{A}$

---

## Concatenatie: $\cdot$

### Concatenatie: $\cdot$

Koppel twee (of meer) reguliere expressies aan elkaar.

### Voorbeeld

- ▶  $a \cdot a \cdot p$ : de taal met de zin 'aap'
- ▶  $0 \cdot 6 \cdot 1 \cdot 1$ : het oude alarmnummer

---

## Concatenatie: $\cdot$

### Concatenatie: $\cdot$

Koppel twee (of meer) reguliere expressies aan elkaar.

### Voorbeeld

- ▶  $a \cdot a \cdot p$ : de taal met de zin 'aap'
- ▶  $0 \cdot 6 \cdot 1 \cdot 1$ : het oude alarmnummer

---

Keuze: |

Keuze: |

Geef een keuze tussen twee (of meer) reguliere expressies aan: óf de ene reguliere expressie óf de andere reguliere expressie wordt 'gevolgd'

Voorbeeld

- ▶  $a|b$ : 'a' of 'b', maar niet beide
- ▶  $(a \cdot b)|c$  'ab' of 'c'

---

Keuze: |

Keuze: |

Geef een keuze tussen twee (of meer) reguliere expressies aan: óf de ene reguliere expressie óf de andere reguliere expressie wordt 'gevolgd'

Voorbeeld

- ▶  $a|b$ : 'a' of 'b', maar niet beide
- ▶  $(a \cdot b)|c$  'ab' of 'c'

---

## Herhaling: \*

### Herhaling: \*

Herhaal een reguliere expressie **nul** of meer keer

### Voorbeeld

- ▶  $a^*$ : '', 'a', 'aa', 'aaa', ...
- ▶  $(a|b)^*$ : '', 'a', 'b', 'aa', 'ab', 'aaaaaaaaabaaaa',  
'babbabbbabbbbaabb', ...



---

## Herhaling: \*

### Herhaling: \*

Herhaal een reguliere expressie **nul** of meer keer

### Voorbeeld

- ▶  $a^*$ : '', 'a', 'aa', 'aaa', ...
- ▶  $(a|b)^*$ : '', 'a', 'b', 'aa', 'ab', 'aaaaaaaaabaaaa',  
'babbabbbabbbbaabb', ...

---

## Groepering: (, )

### Groepering: (, )

Haakjes zoals in de wiskunde: hiermee bouw je ingewikkelder reguliere expressies op.

### Voorbeeld

- ▶  $(a|b)^*$ : '', 'a', 'b', 'aa', 'ab', 'aaaaaaaaabaaaa', 'babbabbbabbbbaabb', ...
- ▶  $(1|2|3|4|5|6|7|8|9|10) \cdot (0|1|2|3|4|5|6|7|8|9)^*$  : gehele getallen

---

## Groepering: (, )

### Groepering: (, )

Haakjes zoals in de wiskunde: hiermee bouw je ingewikkelder reguliere expressies op.

### Voorbeeld

- ▶  $(a|b)^*$ : '', 'a', 'b', 'aa', 'ab', 'aaaaaaaaabaaaa', 'babbabbbabbbbaabb', ...
- ▶  $(1|2|3|4|5|6|7|8|9|10) \cdot (0|1|2|3|4|5|6|7|8|9)^*$  : gehele getallen