



# Regular expressions

Om MacGyver själv får välja!



# Översikt

- Bygger på teori från 50-talet
- Hittar text enligt regler
- "ord som på börjar på a"
- "datum på formatet yyyy-mm-dd"
- Blev populärt med UNIX-verktyg
- Sökning, validering, parsning, filter



# Var stöter man på regexpar?

- Hantering av indata från användare
  - Validering på server/klient
  - Städning
- Tolka url:er, ex. i MVC-router
- Sökning i loggar och databaser
- Filter, ex. omformatering av valuta/datum



# Syntax - översikt

- En regexp är en textsträng
- Slaskar (/) används ofta för att avgränsa
- Bokstäver matchar sig själva
- Metakaraktärer påverkar regexpen
  - `. * | [ ] { } ^ $ \ ? ( )`

# Syntax - bokstäver

abbas sill

- /a/
- /abba/
- /sill/
- /beatles/



# Vilken regexp matchar?

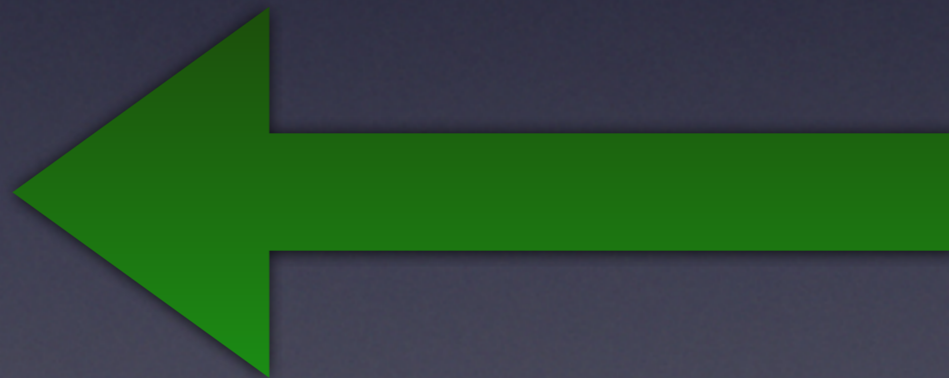
vulkanutbrott

1 /träd/

2 /kant/

3 /nutb/

4 /erik/



# Metakaraktärer

•

- Punkt matchar nästan vad som helst



# Vilken regexp matchar?

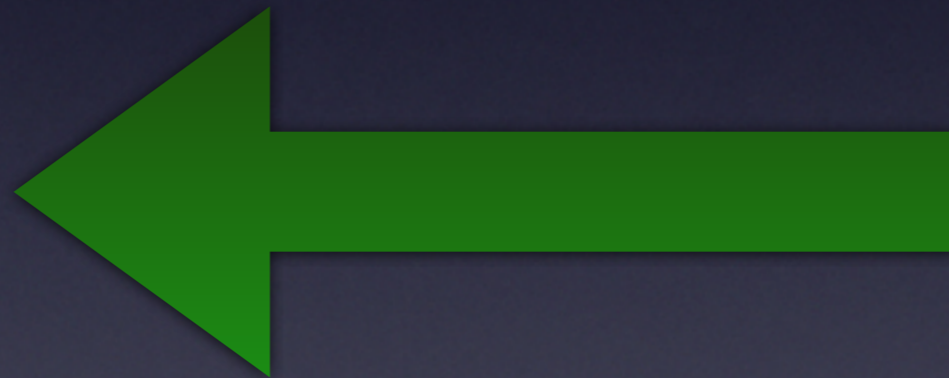
vulkanutbrott

1 /v.k/

2 /...t/

3 /nu..b/

4 /.erik/





# Metakaraktärer

|

- Pipe betyder *eller*
- Avdelar två regexpar

# Vilken regexp matchar?

vulkanutbrott

- 1 /vull|kan/
- 2 /to|na/
- 3 /bron|klan/
- 4 /erik|leander/





# Metakaraktärer

^ \$

- Matchar början och slutet av en textsträng

# Vilken regexp matchar?

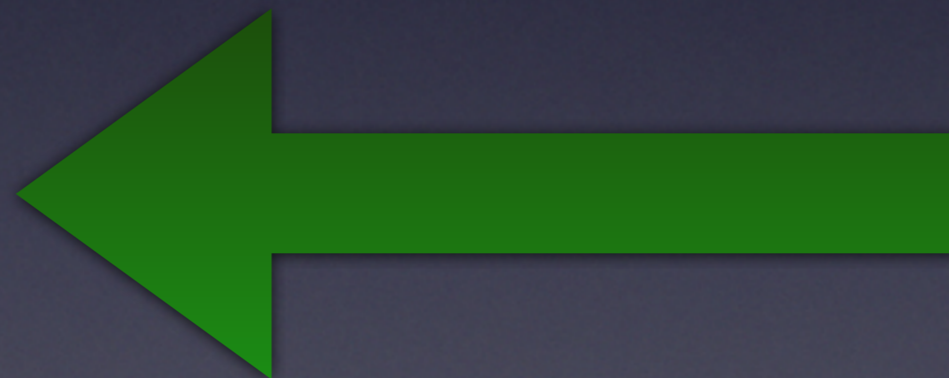
vulkanutbrott

1 /<sup>^</sup>utbrott/


2 /vulkan\$/

3 /brott\$/

4 /<sup>^</sup>erik\$/





A man with long, light-colored hair, wearing dark sunglasses, a red shirt, and a black jacket, is looking towards the camera. He is holding a white model of a rocket or missile. The background is a clear blue sky. A speech bubble is overlaid on the image.

Nu är det  
paus!

# Metakaraktärer

[ ] [ ^ ]

- [ abc ] matchar a, b eller c
- [ ^ abc ] matchar något som inte är a, b eller c



# Vilken regexp matchar?

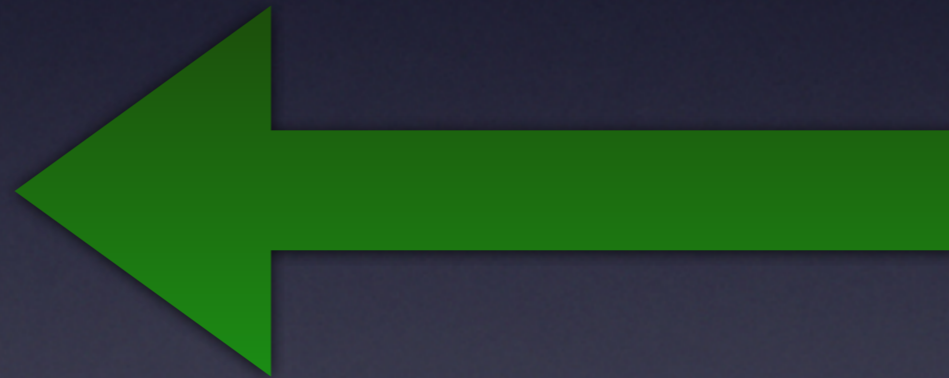
vulkanutbrott

1 `/[a-z]burk/`

2 `/u[ult]b/`

3 `/v[^ulkan]/`

4 `/[er][ik]/`



# Kvantifikatorer

- Hittills har vi matchat en karaktär i taget
- Med kvantifikatorer kan man ange ett antal

? \* + {3} {1,3}



# Kvantifikatorer

- $a?$  matchar noll eller ett  $a$
- $a^*$  matchar noll eller flera  $a$
- $a^+$  matchar en eller flera  $a$
- $a\{3\}$  matchar exakt tre  $a$
- $a\{1,3\}$  matchar minst ett, högst tre  $a$

# Vilken regexp matchar?

vulkanutbrott

1 /u?kan/

2 /t?t?au/

3 /vulkan?/

4 /erik?/





# Vilken regexp matchar?

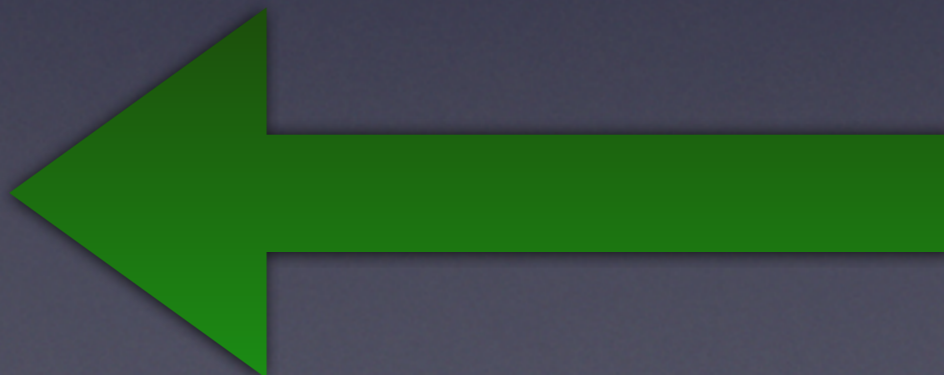
vulkanutbrott

1 /a\*aa\*a/

2 /utb\*t/

3 /ot\*o/

4 /e\*r\*i\*k/



# Vilken regexp matchar?

vulkanutbrott

1 /t+o/

2 /an+u+/  


3 /tt+t/

4 /erik+idun/



# Vilken regexp matchar?

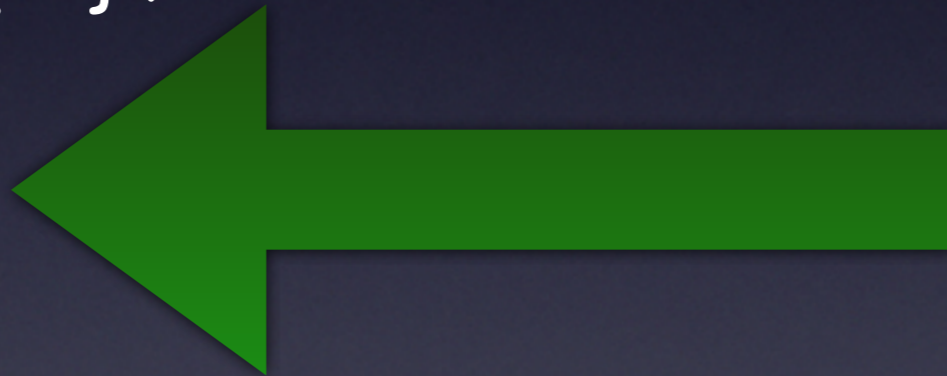
vulkanutbrott

1 /t{1,2}b{2}/

2 /rot{2}/

3 /ka{2,}/

4 /erik{1337}/



# Vilken sträng matchar?

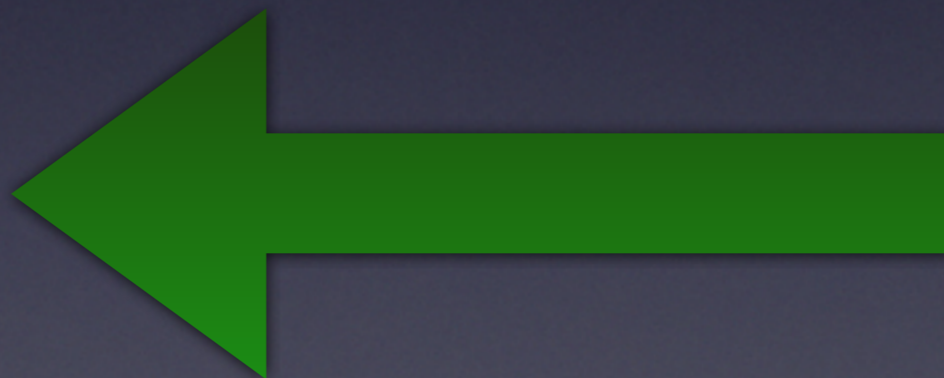
```
/^[ab]b{2}a?$/
```

1 abbas sill

2 ab

3 abba

4 erik





# Vilken sträng matchar?

```
/^[a-z]+\.[a-z]+@macrohard\.com$/
```

- 1 mikael@macrohard.com
- 2 pontus.walck@macrohard.com
- 3 m.z.tylmad@macrohard.com
- 4 erik.leander@microsoft.com



# Grupper

- Parenteser delar upp regexpar i grupper
- Grupper kan man komma åt som variabler

```
/([a-z]+\.[a-z]+)@macrohard\.[a-z]+$/
```

- Grupp ett innehåller personen på macrohard
- Grupp två innehåller domän-ändelsen



# Regexpar i PHP

- Två vanliga funktioner
  - `preg_match`
  - `preg_replace`

```
<?php
```

```
    if (preg_match('/MacGyver/', $text))  
        print 'coolt!';
```

```
?>
```



```
$adminsPattern = '/^mikael$|^pontus$/';  
if (preg_match($adminsPattern, $_SESSION["user"]))  
    $isAdmin = true;
```

```
$userPagePattern = '/^user\/([0-9]+)$/';  
$dogPagePattern = '/^dog\/([0-9]+)$/';
```

```
if (preg_match($userPagePattern,  
              $_SERVER['QUERY_STRING'],  
              $matches)  
    renderUserPage($matches[1]);
```


```
else if (preg_match($dogPagePattern,  
                   $_SERVER['QUERY_STRING'],  
                   $matches)  
    renderDogPage($matches[1]);
```

```
<?php
```

```
    $XSScleanText = preg_replace('/</', '&lt;', $text);
```

```
?>
```



A man with brown hair, wearing a brown leather jacket, is leaning on the side of a white car. He is smiling and looking towards the camera. The background shows a desert landscape with hills under a clear blue sky. A speech bubble is overlaid on the image, containing the text "Nu är det slut!".

Nu är det  
slut!