

# Regular Expressions in JavaScript

- Besondere Zeichen
- Optional, Min, Max,...
- Zeichenklasse
- Beispiele
  - Zeichenklassen & Wiederholungen
  - Wordboundary
  - UTF8/16/...
  - Miscellaneous
    - Zeichenklassen
    - Oder
    - Within single quotes
    - n'te Gruppe (naive Annahme: Zwischen Quotes)
    - n'te Gruppe (besser)

# Besondere Zeichen

`^ $ . , - * + ? = ! : | ( ) [ ] { } \`

|                       |   |
|-----------------------|---|
| <code>\f \n</code>    | Seitenvorschub, Zeilenumbruch           |
| <code>\r \t \v</code> | Wagenrücklauf, H-Tabulator, V-Tabulator |
| <code>\d \D</code>    | digit, non digit                        |
| <code>\s \S</code>    | white space, non white space            |
| <code>\w \W</code>    | word, non word                          |
| <code>\b \B</code>    | word boundary, non word boundary        |
| <code>\cI</code>      | horizontal tab                          |
| <code>\p{L}</code>    | any kind of letter from any language    |

## Optional, Min, Max,...

X is a placeholder for any regular expression (own notation)

|                     |                            |
|---------------------|----------------------------|
| <code>X?</code>     | optional                   |
| <code>X*</code>     | 0 oder mehr                |
| <code>X+</code>     | 1 oder mehr                |
| <code>X{n}</code>   | n mal                      |
| <code>X{n,}</code>  | mindestens n mal           |
| <code>X{m,n}</code> | mindestens m mal max n mal |

## Zeichenklasse

|                    |                            |
|--------------------|----------------------------|
| <code>[abc]</code> | Zeichenklasse: eines davon |
|--------------------|----------------------------|

# Beispiele

## Zeichenklassen & Wiederholungen

|                         |                        |
|-------------------------|------------------------|
| <code>[^0-9]</code>     | Keine Ziffern          |
| <code>[^0-9]+\$</code>  | Am Ende keine Ziffern  |
| <code>[0-9]{4,6}</code> | Vier bis sechs Stellen |

## Wordboundary

|                  |                                   |
|------------------|-----------------------------------|
| <code>e\b</code> | jedes e vor einem "Wordboundary". |
|------------------|-----------------------------------|

Schule Haus Hose Matratze

## UTF8/16/...

`\u{1f310}`

## Miscellaneous

### Zeichenklassen

|                          |                         |
|--------------------------|-------------------------|
| <code>[^0-9]+\$</code>   | Keine Ziffern am Ende   |
| <code>[1-9][0-9]*</code> | Zahl ohne führende Null |

## Oder

|                       |               |
|-----------------------|---------------|
| <code>http ftp</code> | http oder ftp |
|-----------------------|---------------|

## Within single quotes

```
'([^\']*)*'
```

Was heißt "hot", ist es 'heiß' oder 'kalt'?

## n'te Gruppe (naive Annahme: Zwischen Quotes)

Alles zwischen double- or single quotes. Der Fehler: Text dazwischen wird auch ausgewählt:

```
(['"]).*\1
```

Was heißt "hot", ist es 'heiß' oder 'kalt'?

## n'te Gruppe (besser)

Alles zwischen double- or single quotes. (Korrigierte Version des obigen Beispiels)

```
(['"])([^\1"])*\1
```

Was heißt "hot", ist es 'heiß' oder 'kalt'?