

InDesign

Suchen

GREP

Reguläre Ausdrücke

Fundstellen

Ersetzen

Zeichenklassen

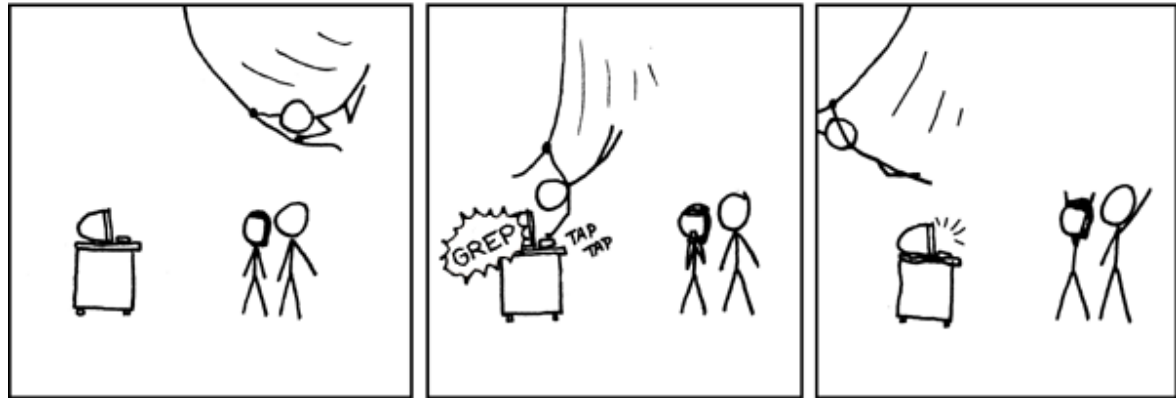
GREP in InDesign

28. April 2012

Deutsche Publishing Konferenz

Kontakt: gregor.fellenz@publishingx.de

Folien: <http://www.publishingx.de/dokumente>



Lizenz: © ⓘ ⓘ | Quelle: Remixed from <http://xkcd.com/208> | Autor: Randall Munroe

Schöner suchen und ersetzen

InDesign bietet ab CS3 die Suche mit **Regulären Ausdrücken** bzw. **GREP**.

- Variable **Suche** nach Zeichenklassen (Ziffern, Leerräume, Großbuchstaben etc.), Wiederholungen und Positionen.
- Bei der Suche gefundene Stellen in der **Ersetzung** wiederverwenden.
- Seit InDesign CS4 **GREP-Stile**.
- Finden Sie die Suchanfragen unübersichtlich und komplex?
Abstraktion ermöglicht erst die Mächtigkeit der Suchausdrücke.

Beispiele

- Ein Wort ist falsch geschrieben und zwar auf 3 unterschiedliche Arten.

`F[als]ch`

- Seitenverweise in der Form `(Seite 100)` in die Form `→ Seite 100` bringen.

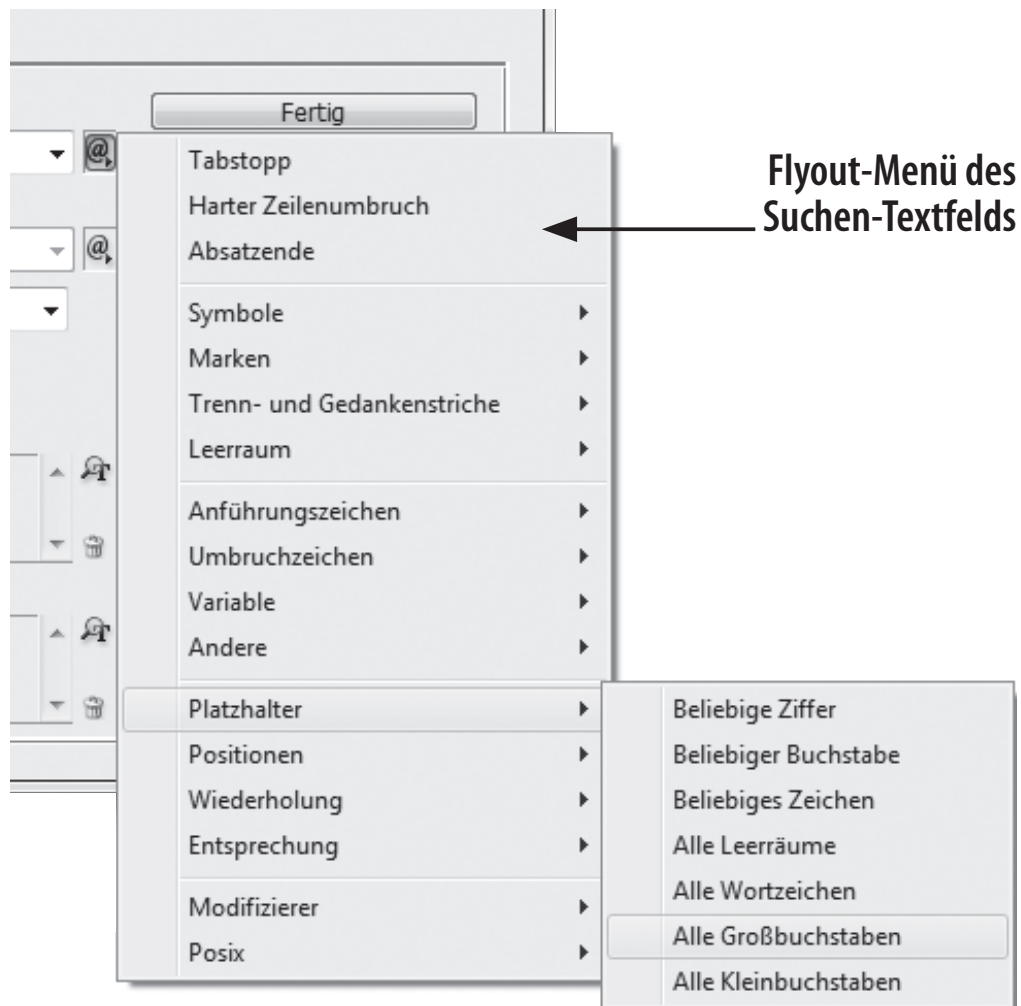
`\(Seite (\d+)\)` Ersetzen durch `→ Seite $1`

- Alle Kapitel eines Buchs müssen mit einem Absatzformat formatiert werden. Sie beginnen imm mit `Kapitel`

`^Kapitel.*?` Ersetzen durch `$0` und Formateinstellung

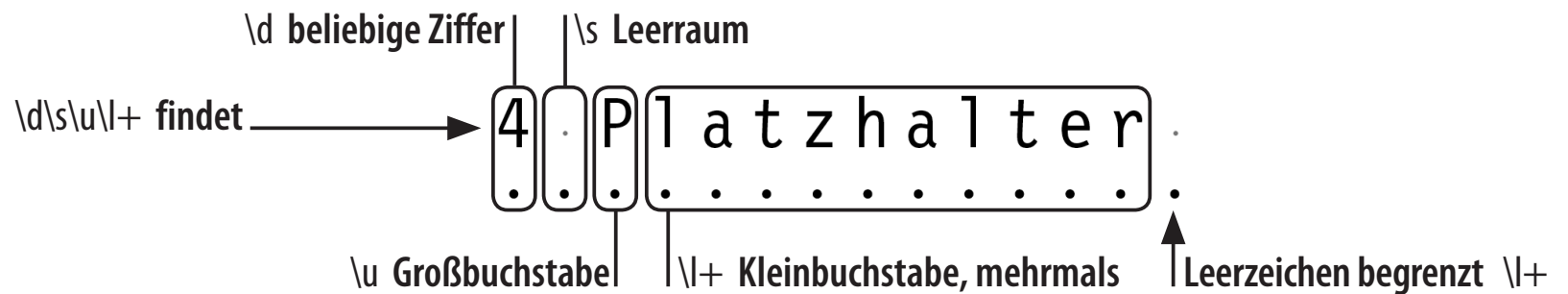
- Eine Zahlenreihe enthält am Ende Kommas, Kommas entfernen (ist `3,4`, soll `3,4`)

Suche: `(\d+,\d+)`, Ersetzen durch `$1`



Die Suche mit Regulären Ausdrücken

- Zeichen mit **spezieller Bedeutung**
 - z.B. `~S` oder `.`
- Maskierung/Reservierte Zeichen
`\ . * + ? () { } [^ $ | ~`
- Durch Steuerzeichen verursachte »Probleme« sind der einzige Grund für die normale Suche



Platzhalter und Zeichenklassen

- Der wichtigste und einfachste Platzhalter ist der Punkt `.`. Er findet jedes beliebige Zeichen, mit Ausnahme vom Absatzende.
 - Mit `..` findet man zwei beliebige, aufeinanderfolgende Zeichen.

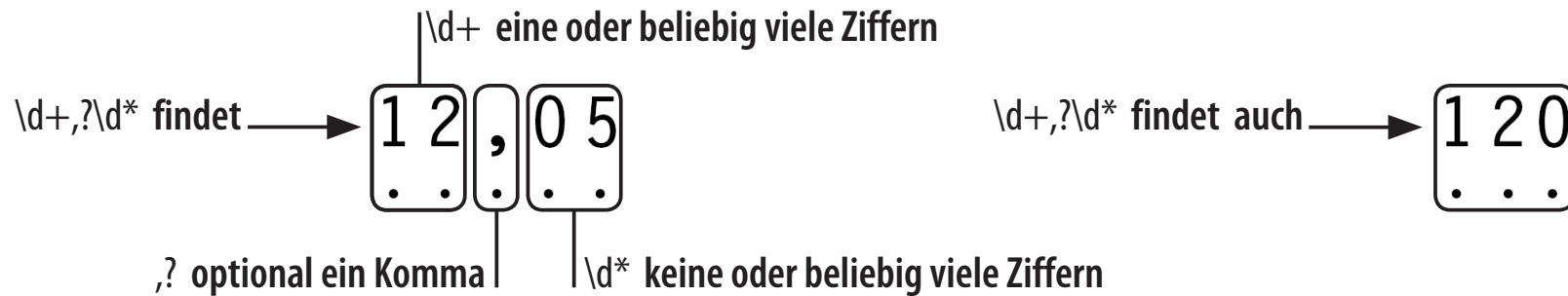
Platzhalter repräsentieren eine bestimmte Menge an Zeichen, sie werden auch als **Zeichenklassen** bezeichnet.

- Mit `\d` findet man alle Ziffern.
- Mit `\u` findet man einen beliebigen Großbuchstaben, mit `\l` einen Kleinbuchstaben.
- Mit `\s` findet man beliebige Leerräume und Absatzumbrüche.

Übersicht Zeichenklassen

| Zeichenklasse | Beschreibung |
|---------------|---|
| . | Ein beliebiges Zeichen aber nicht das Absatzende (Return, etc.) oder einen erzwungenen Zeilenumbruch (Soft-Return). |
| \u | Alle Großbuchstaben (engl. uppercase) |
| \l | Alle Kleinbuchstaben (engl. lowercase) |
| [\u\l] | Beliebiger Buchstabe |
| \w | Alle Wortzeichen (inklusive Zahlen und Unterstrich _ ohne Bindestrich -) |
| \d | Alle Ziffern (engl. digits) |
| \D | Alles außer Ziffern. |
| \s | Alle Weißräume. Also alle Wortabstände, Tabulatoren und Umbruchzeichen. (engl. space) |
| \S | Wieder genau das Gegenteil, also keine Weißräume. |
| \p{Zs} | Alle Leerräume (Sonderfall alle Leerräume mit Tabulator (\p{Zs} \t)) |

Wenn ein Großbuchstabe verwendet wird, wird genau das Gegenteil gefunden. **\D** findet alles außer Ziffern.



Wiederholungen

- Mit dem Plus-Zeichen `+` mehrere Zeichen vom gleichen Typ suchen.
 - `.+` Alle Zeichen bis zum Absatzende
 - Mit `M\1+` entsprechend alle Worte, die mit einem großen M beginnen und denen Kleinbuchstaben folgen: `Max`, `Moritz`, `Montag`, `Mars` usw. aber nicht `MacOS`.
- Mit dem Sternchen `*` kein oder mehrere Zeichen Suchen.
 - `\s*` für optionale Leerräume.
- GREP sucht immer so viel wie möglich, es ist **gierig**. Mit `?` nur den kürzesten Treffer.

Übersicht Wiederholungszeichen

| Wiederholungen | Beschreibung |
|--------------------|--|
| <code>?</code> | Null oder ein Mal |
| <code>*</code> | Null oder mehrere Male |
| <code>+</code> | Ein oder mehrere Mal |
| <code>??</code> | Null oder ein Mal (kürzeste Entsprechung/non greedy) |
| <code>*?</code> | Null oder mehrere Male (kürzeste Entsprechung/non greedy) |
| <code>+?</code> | Ein oder mehrere Mal (kürzeste Entsprechung/non greedy) |
| <code>{n}</code> | Der Suchbegriff muss exakt n-mal vorkommen. |
| <code>{n,m}</code> | Der Suchbegriff Ausdruck muss mindestens n-mal und darf maximal m-mal vorkommen. |
| <code>{n,}</code> | Der Suchbegriff Ausdruck muss mindestens n-mal und darf öfter vorkommen. |

Praxisbeispiel Preisliste

Eine Liste finden Sie in der Datei `Beispiel_Preisliste.indd`

! Entwickeln Sie einen Regulären Ausdruck, der die verschieden formatierten Preisangaben in der Preisliste findet!

! Lösung: `.+?\s+\d+[\.,]? \d{0,2} \s* (EUR|€)`

Besonders günstige Textsuche 1 EUR

Wiederholungen kosten mehr 5900 EUR

Zeichenklassen finden für nur 2EUR

Mehr Raum 3 EUR

Suche mit Varianten 4 €

Nachkommastellen 100,10 EUR

Mit Punkt und Komma 150.99 EUR

Positionsangaben 200,00 EUR

~~Suche~~ Ersetze

Intelligenter ersetzen

Reguläre Ausdrücke bieten die Möglichkeit, das Suchergebnis bei der Ersetzung weiterzuverwenden. Der Zugriff auf die Fundstellen geschieht mit so genannten Rückwärtsreferenzen z.B. `$0` für den gesamten Suchtreffer.

- Vollständige Fundstelle in der Ersetzung verwenden

Suche nach E-Mail: `[\w.]+?@[\w.]+?`

Ersetzung durch: `E-Mail: $0`

- Bestimmte Bereiche mit Klammern `()` zusammenfassen

■ Beispiel Preisliste

Bringen Sie die Liste in das Format: `Preis Tabulator EUR Tabulator Listeneintrag`

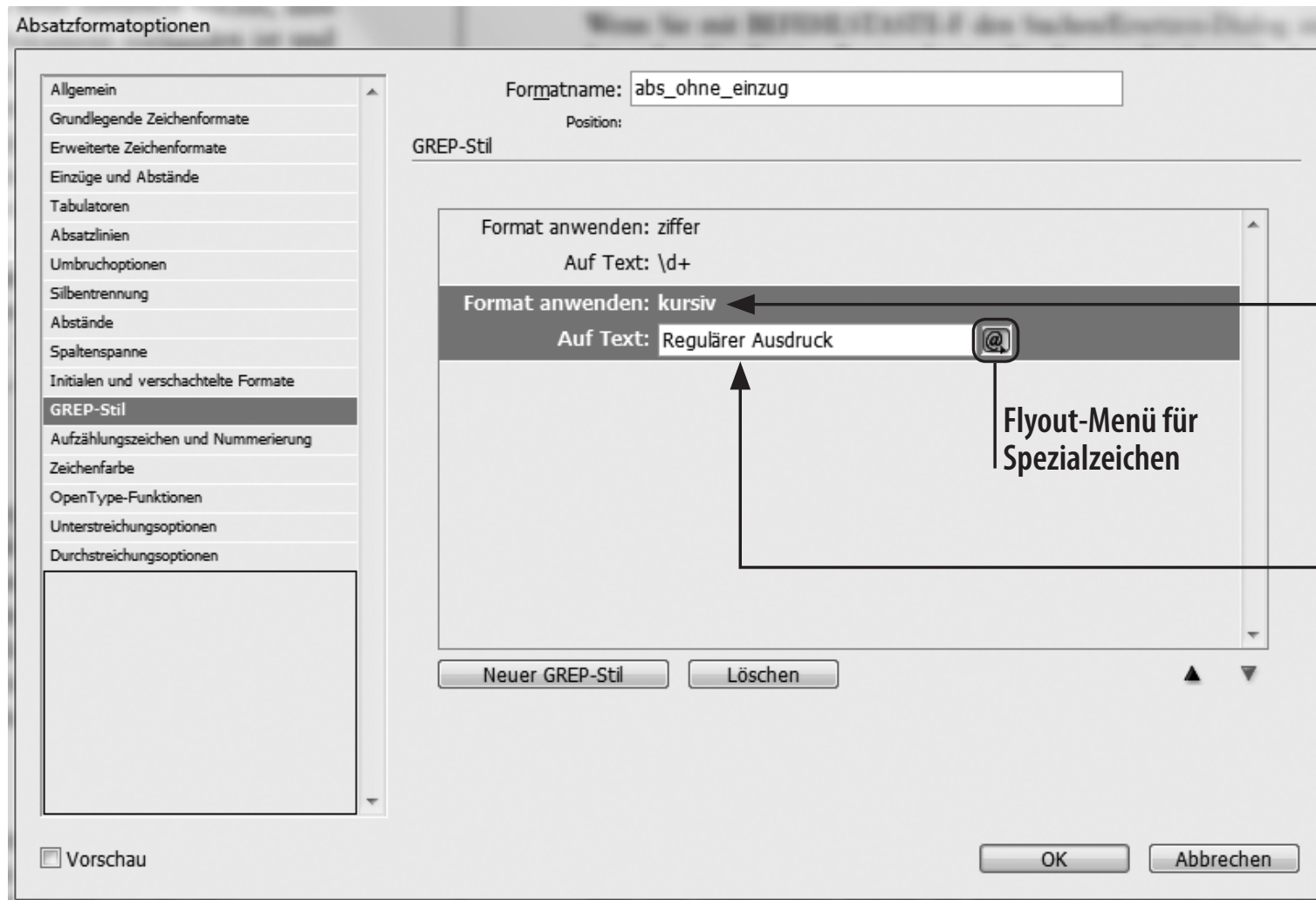
Suche: `(.+?)\s+(\d+[.,]?[0,2])\s*(EUR|€)`

Ersetze: `$2\tEUR\t$1`

- Maskierung/Reservierte Zeichen bei der Ersetzung

`\ $ ~`

Da gab's doch noch... GREP-Stile



Zeichenformat, mit dem
der Reguläre Ausdruck
formatiert wird

Flyout-Menü für
Spezialzeichen

Beliebiger
Regulärer Ausdruck

In eigener Sache

InDesign automatisieren – Skripting, GREP & Co.

Das Buch zur InDesign Automation mit einer Skripting Einführung und vielen Praxistipps zu EPUB, XML und GREP.

Auf der Homepage zum Buch <http://www.indd-skript.de> gibt es Leseproben und alle Beispiele

- Klassisch auf Papier
ISBN: 978-3-89864-734-2
Preis: 34,90 Euro (D), 35,90 Euro (A)
- EPUB
ISBN: 978-3-89864-882-0
Preis: 27,90 EUR



Vielen Dank für Ihre Aufmerksamkeit!

Fragen und Anregungen?

Die Folien zum Download:
<http://www.publishingx.de/dokumente>

E-Mail:
gregor.fellenz@publishingx.de