

InDesign

Suchen

GREP

Reguläre Ausdrücke

Fundstellen

Zeichenklassen

Ersetzen

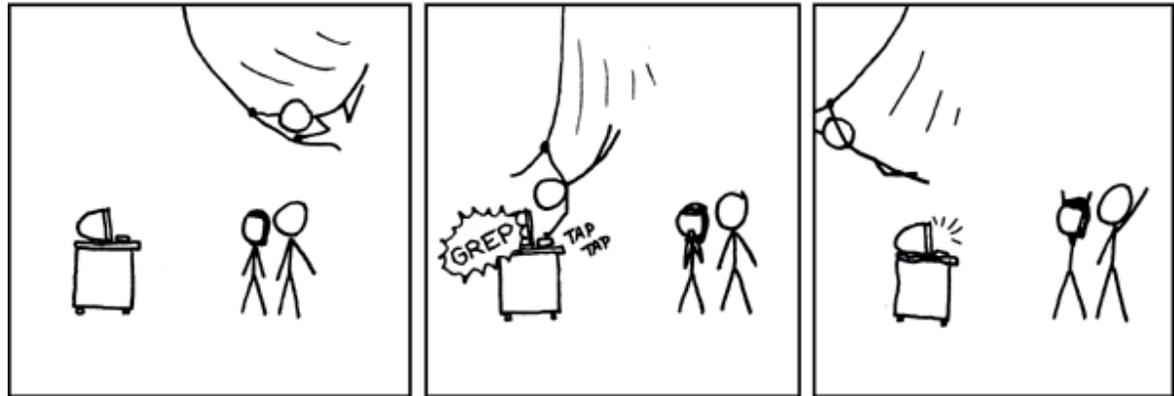
GREP in InDesign

5. November 2015
#swisspublishingdays

Kontakt: gregor.fellenz@publishingx.de
Folien: <http://www.publishingx.de/dokumente>

Übersicht

- Suchen und Ersetzen mit GREP
- GREP-Tools
- Beispiele/Wünsch dir was!



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**.

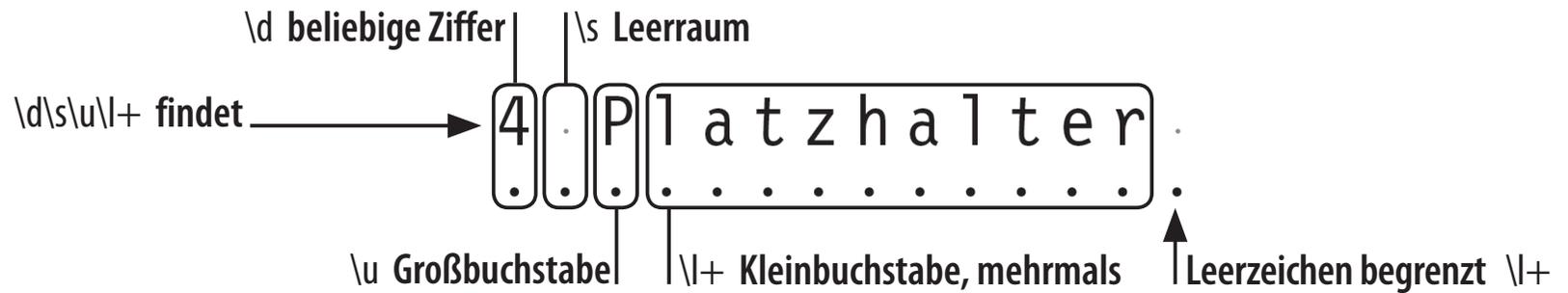
Variable/unscharfe **Suche** nach Zeichen (z.B. Ziffern, Leerräume, Großbuchstaben etc.), Wiederholungen, Positionen und Alternativen.

Bei der **Ersetzung** in der Suche gefundene Stellen wiederverwenden.

Seit InDesign CS4 **GREP-Stile**.

Beispiele

- **Alle Ziffern im Dokument:** `\d`
Ziffern in kleinerer Schrift oder Überprüfung von Einheiten.
- **Alle Leerräume im Dokument:** `\h`
Die doppelten Leerzeichen...
- **Zeichenauswahl** `[a-n]`
Lokalisierung von bestimmten Kombinationen z.B. Produktnummern.
- **Variationen** `(Müller|Mue1ler)`
Ein Name ist in unterschiedlich geschrieben worden.
- **Mehr als nur einmal** `+`
`\d+` oder `\h+`
- **Komplexe Ersetzungen** `Seite (\d+)` Ersetzen durch `→ $1`
Seitenverweise in der Form `Seite 100` in die Form `→ 100` bringen.



Platzhalter/Joker

- Der einfachste **Platzhalter** ist der Punkt `.`.
Er findet jedes beliebige Zeichen, mit Ausnahme von Umbrüchen.
Mit `..` findet man zwei aufeinanderfolgende Zeichen.

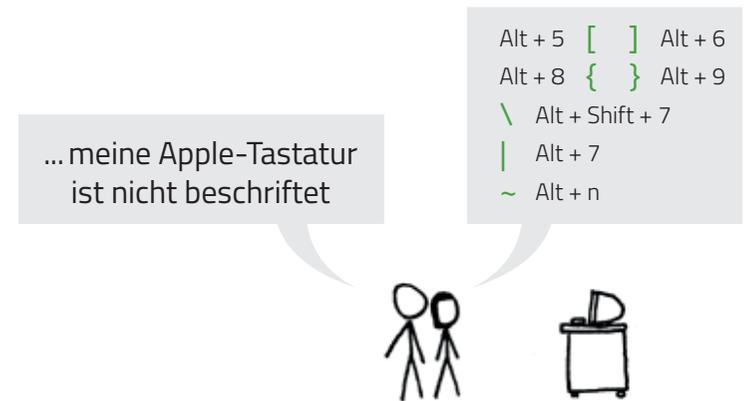
Platzhalter repräsentieren eine bestimmte Menge an Zeichen, sie werden **Zeichenklassen** genannt.

- Bereits bekannt: Mit `\d` alle Ziffern lokalisieren
Mit `\u` findet man einen beliebigen Großbuchstaben, mit `\l` einen Kleinbuchstaben.

Zeichen- klasse	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.
\h	Alle Leerräume (horizontal space; kein Tab rechte Ausrichtung)
(\h ~y)	Sonderfall alle Leerräume mit Tabulator rechte Ausrichtung

Wenn ein Großbuchstabe verwendet wird, [wird genau das Gegenteil gefunden.

\D findet alles außer Ziffern.



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

Beispiel

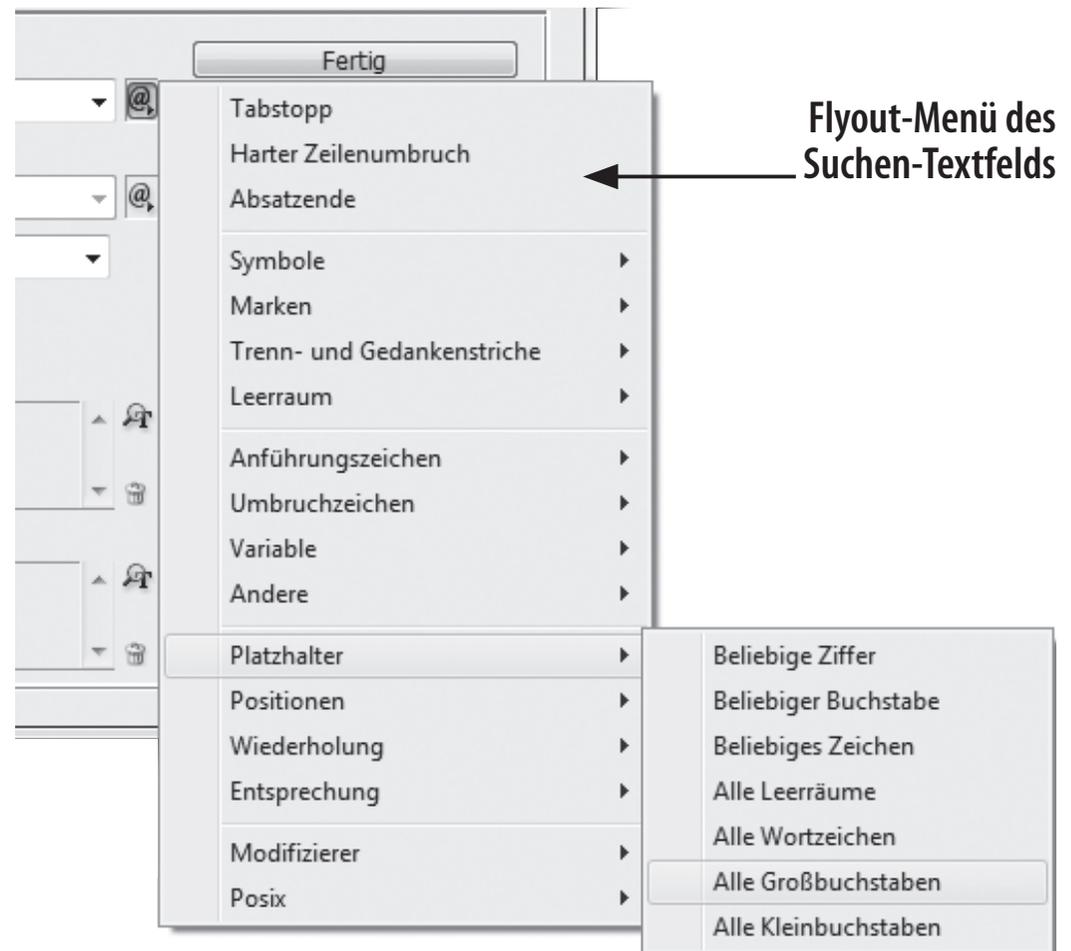
Eine Beispieldatei finden Sie in der Datei `01_Leerraum.indd`

Mehrfache Leerräume ersetzen:

- **Schlechte Lösung:** `\s+`
Trifft auch Umbruchzeichen.
- **Bessere Lösung:** `\h+`
Spezialzeichen (Indexmarken, Notizen, XML-Tags ...) werden ggfs. gelöscht.
- **Noch Besser:** `\h(?:=\h)`
Das `(?:=\h)` ist ein Lookahead (noch nicht vorgestellt)

Zeichen mit spezieller Bedeutung

- z.B. \ oder .
- Maskierung/Reservierte Zeichen
\\ . * + ? () { } [^ \$ | ~
- Wenn diese konkret gefunden werden sollen, muss ein \ vorangestellt werden (Maskierung).
- Durch Steuerzeichen verursachte »Probleme« sind der einzige Grund für die normale Suche



Zeichenauswahl

- Mit eckigen Klammern `[]` beliebige Auswahl festlegen
`S[kc]ript` findet die Variationen `Skript` und `Script`
- **Maskierung** – In der Auswahl reservierte Zeichen:
`] ^ - \ ~`
- `[gjpgy]` alle Buchstaben mit Unterlängen (evtl. zusätzlich *ß* und *f*)
z.B. Unterstreichungen aufheben.
- `200[5-9]` alle Jahreszahlen zwischen 2005 und 2009
- `[\u\1\~]+` praktisch alle Wörter: Groß-, Kleinbuchstaben und den Bindestrich
- **Umkehrung** mit `[^a-n]` Findet alles außer a-n in Kleinbuchstaben.

Sonderzeichen

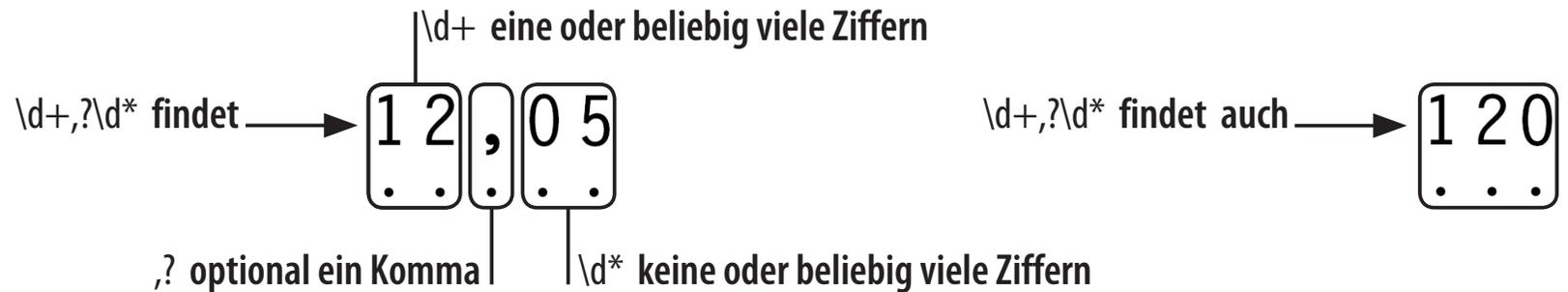
Viele Zeichen, wie z.B. die Zeilenschaltungen, können nicht direkt in die Such- bzw. Ersetzungsanfrage eingegeben werden.

- Standardsonderzeichen:

- `\r` (Umbrüche außer Harter Zeilenumbruch) `\n` (Harter Zeilenumbruch) `\v` Alle Umbrüche
 - `\t` (Tabulator)

- Alle anderen werden mit der Tilde `~` erzeugt – z.B. `~S` für ein geschütztes Leerzeichen

Zeichen	Beschreibung
<code>\t</code>	Tabulator
<code>\r</code>	Absatzende bzw. Zeilenschaltung (Return) aber auch Spaltenumbruch, Seitenumbruch etc. sowie erzwungener Zeilenwechsel (Soft-Return).
<code>\n</code>	Erzwungener Zeilenwechsel (Soft-Return)
<code>~S</code>	Geschütztes Leerzeichen
<code>~></code>	Halbgeviert
<code>~<</code>	Achtelgeviert
<code>~=</code>	Halbgeviertstrich, Gedankenstrich
<code>~-</code>	Bedingter Trennstrich
<code>~a</code>	Marker für verankerte Objekte
<code>\~</code>	Tilde



Wiederholungen

Das Wiederholungszeichen, auch Quantifizierer genannt, bezieht sich immer auf das vorhergehende Zeichen, eine Zeichenklasse oder das vorhergehende Wort in runden Klammern – Er kann nicht alleine stehen.

- Wiederholungszeichen **?** (Im Menü **Null oder ein Mal**)
 Mit `dass?` z.B. alle Vorkommen von `das` und `dass`
 Mit `(Max)? Mustermann` jede Erwähnung von Herrn Mustermann – mit und ohne Vorname
- Wiederholungszeichen **+** (Im Menü **Ein oder mehrere Male**)
`+` mehrere Leerzeichen
`[12][A-N]\d\d\h.+` findet man `1G78 Staubsauger` oder `1H32 Wischmops`
- Wiederholungszeichen ***** (Im Menü **Null oder mehrere Male** korrekt: Null oder beliebig oft)
`\d+,?\d*` findet sowohl `15` als auch `14,9999`



- GREP sucht immer so viel wie möglich, es ist **gierig**. Mit `?` nur den kürzesten Treffer.
 - Mit `.+?\.` wird nur der erste Satz eines Absatzes gefunden
- **Genauere Anzahl der Vorkommen festlegen** Zwei- oder dreistellige Zahlen mit `\d{2,3}`

Wiederholungen	Beschreibung
<code>?</code>	Null oder ein Mal
<code>*</code>	Null oder beliebig oft
<code>+</code>	Ein oder mehrere Mal
<code>??</code>	Null oder ein Mal (kürzeste Entsprechung/non greedy)
<code>*?</code>	Null oder beliebig oft (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 `02_preisliste.indd`

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

! Lösung: `.+?\h+\d+ EUR`

Besonders günstige Textsuche	1 EUR
Wiederholungen kosten mehr	5900 EUR
Zeichenklassen finden für nur	2 EUR
Mehr Raum	3 EUR
Suche mit Varianten	4 EUR
Nachkommastellen	100 EUR
Sind auch wichtig	150 EUR
wurden hier aber weggelassen	200 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.

- **Vollständige Fundstelle** in der Ersetzung verwenden
Suche nach Telefonnummer: `[\d\h+()\./\-]{6,}`
Ersetzung durch: `Te1.: $0`
- Bestimmte Bereiche mit Klammern `()` zusammenfassen
Suche: `(.+?)\h+(\d+)` EUR
Ersetzung: `$1\t$2~SEUR`
- Maskierung/Reservierte Zeichen bei der Ersetzung `\ $ ~`



Lizenz: © ⓘ ⓘ | Autor: Randall Munroe

Ausschau halten

Bei der Verwendung von sogenannten **Look Around Assertions** wird geprüft, ob vor bzw. nach dem eigentlichen Ausdruck ein anderer Regulärer Ausdruck steht. Besonders wichtig für **GREP-Stile**

Lookbehind

- **Positiv** Brüche: Ziffern nach einem Bruchstrich: `(?<=/) \d+`
- **Negativ** Preise vor denen nicht EUR steht: `(?<!EUR) \h+ \d+, \d \d`
- **Achtung:** Lookbehind **nur** mit fester Länge möglich (Wiederholungen nicht einsetzbar)

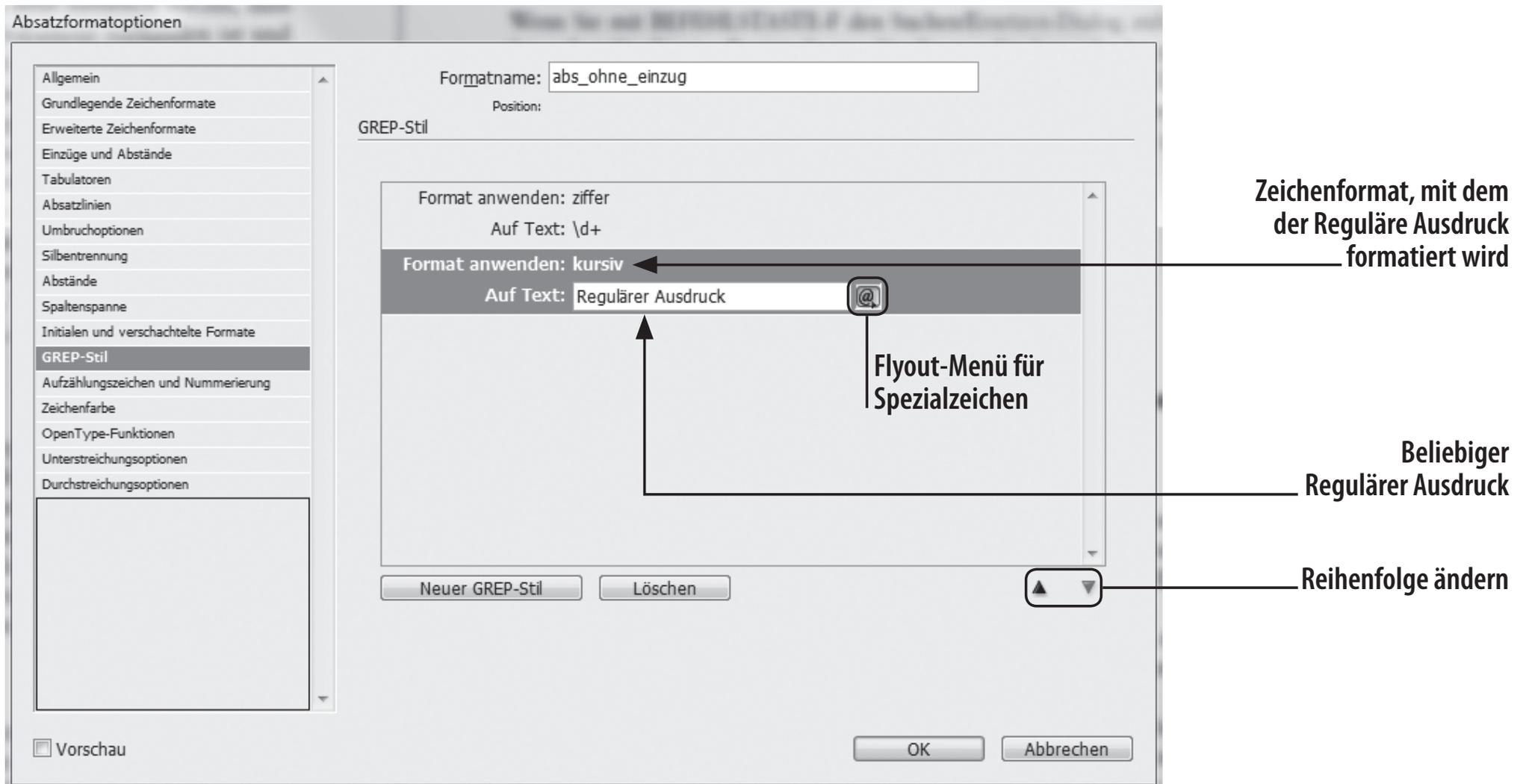
Lookahead

- **Positiv** Ziffern denen die Einheit `cm` folgt: `\d+ (?= \hcm)`
Mehrere Leerräume `\h (?= \h)`
- **Negativ** Geschäftsbericht, AG hinter dem Firmennamen prüfen: `Firma \h+ (?!AG)`

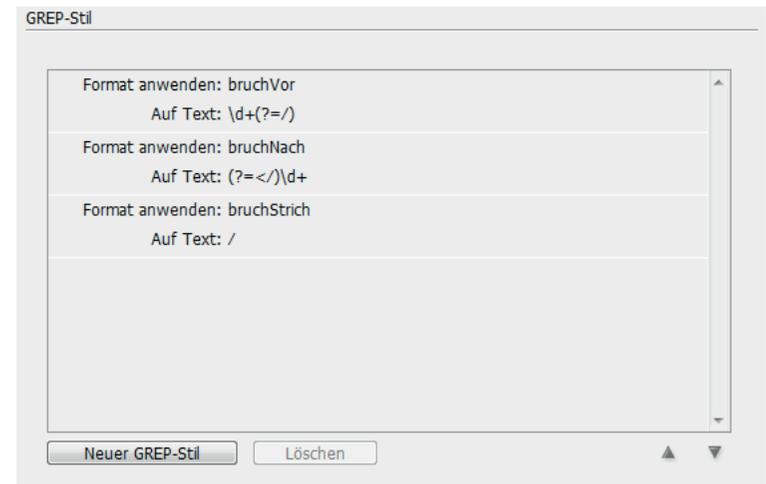
GREP-Stile

Look Around Assertions eignen sich gut für die Auszeichnung von Texten mit GREP-Stilen.

- GREP-Stile überschreiben sich gegenseitig (von oben nach unten)



1²/₇ Ein



Brüche mit einem GREP Stil automatisch formatieren

3/4 soll zu $\frac{3}{4}$ werden

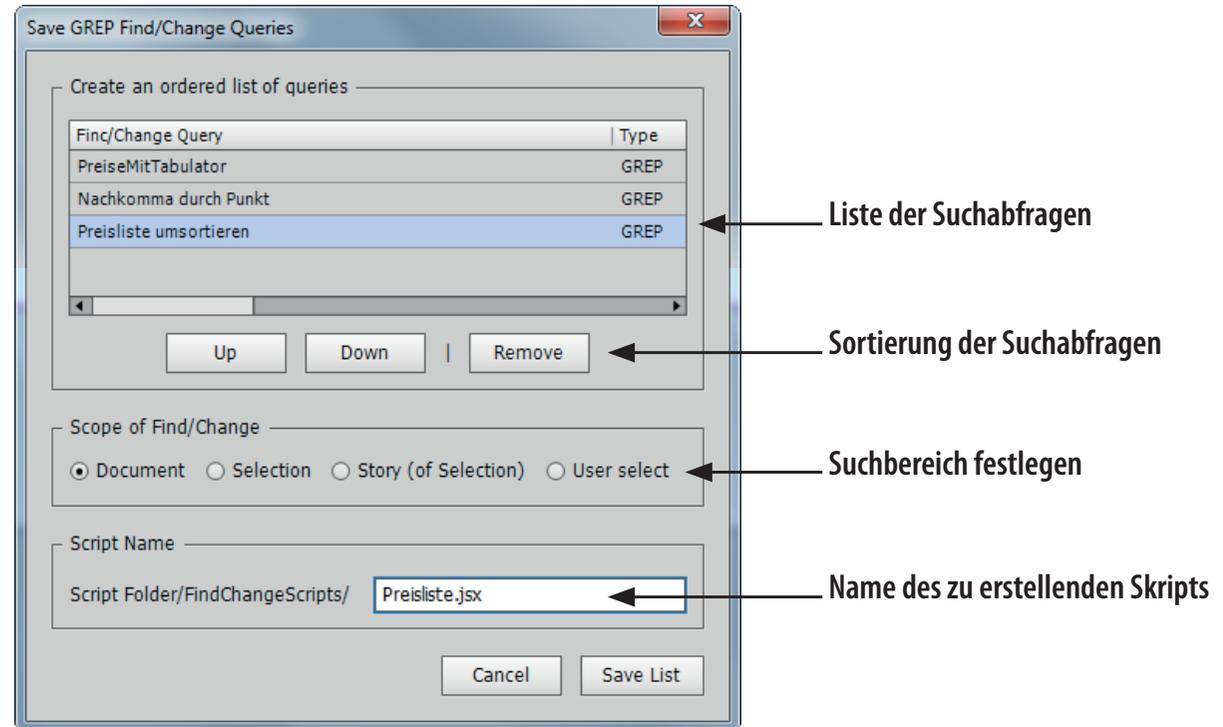
Beispieldokument 03_BruchGrepStil.indd

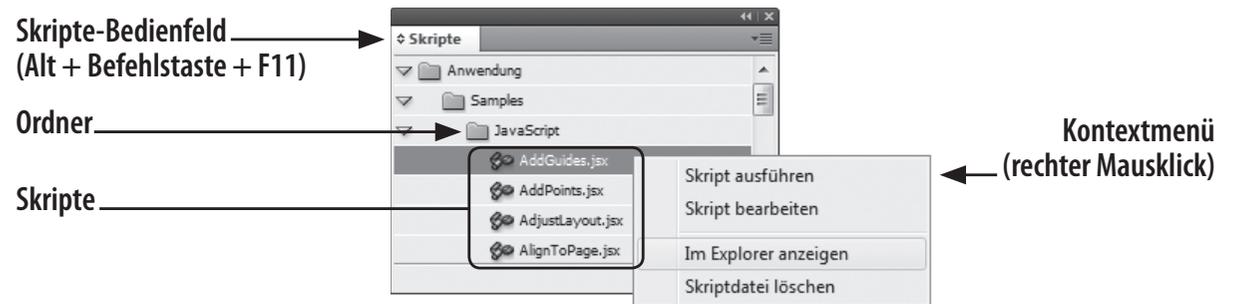
- Ziffer vor Bruchstrich `\d+(?=/)`
- Ziffer nach Bruchstrich `(?=</)\d+`
- Bruchstrich `(?=<\d)/(?=/\d)`

GREP-Tools

ChainGREP.jsx

- GREP-Abfragen im Suchen/Ersetzen-Dialog speichern.
- Skript ChainGREP.jsx aufrufen
<http://www.publishingx.de/download/ChainGREP.jsx>
- Liste der GREP-Abfragen sortieren
- Namen für das zu speichernde Skript eintragen.
- Die Skripte werden im Unterordner FindChangeScripts gespeichert.





Skripte installieren

Dreh- und Angelpunkt für die Verwendung von Skripten in InDesign ist das **Bedienfeld *Skripte***

- CS3/CS4: *Fenster > Automatisierung > Skripten*
- CS5-CC: *Fenster > Hilfsprogramme > Skripte*

Über das Kontextmenü eines Eintrags im Skript-Bedienfeld erreichen Sie den Ordner im Dateisystem.

Im sich öffnenden Explorer/Finder navigieren Sie in den Unterordner **Scripts Panel**. In diesen Ordner kopieren Sie das Skript.

Wechseln Sie nun zurück zu InDesign. Das Skript erscheint im Bedienfeld und kann nun durch einen Doppelklick ausgeführt werden.

Weitere Tools

Die Tool Sammlung von Peter Kahrel finden Sie unter http://www.kahrel.plus.com/indesign/grep_matters.html

Für komplexe Ausdrücke ist der **GREP Editor** zu empfehlen. Mit ihm können Abfragen in mehreren Zeilen – und damit deutlich übersichtlicher – eingegeben werden. Er enthält eine vollständige Liste von Platzhaltern und Steuerzeichen. Die InDesign Suche kann direkt vom GREP Editor gesteuert werden.

Das Skript **GREP Manager** ermöglicht es, mehrere GREP-Abfragen zusammenzustellen und hintereinander ablaufen zu lassen. Man kann dort auch Abläufe für die spätere Verwendung speichern.

Dokumentation: http://www.kahrel.plus.com/indesign/grep_query_manager.html

Peter Kahrel hat übrigens auch das Standardwerk *GREP in InDesign* verfasst, welches in englischer Sprache bei O'Reilly als E-Book erhältlich ist.

Andere Software ...

- **Text Editor SublimeText**

Mac/Windows <http://www.sublimetext.com/>

- **Dateien umbenennen**

Mac/Windows <http://www.publicspace.net/windows/BetterFileRename/index.html>

In eigener Sache

InDesign automatisieren – Skripting, GREP & Co. 2. Auflage

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

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

- Klassisch auf Papier
ISBN: 978-3-86490-235-2
Preis: 36,90 Euro (D)
- EPUB
ISBN: 978-3-86491-684-7
Preis: 29,99 EUR



Vielen Dank für Ihre Aufmerksamkeit!

Fragen und Anregungen?

E-Mail: gregor.fellenz@publishingx.de

Twitter: [grefel](#)

Blog: <http://www.indesignjs.de>