

1 Schöner suchen und ersetzen mit GREP

Bei der Arbeit in InDesign kommt man oft mit der Suchen/Ersetzen-Funktion in Berührung. Viele Anwender beschränken sich auf die Suche nach bestimmten Texten, die gegebenenfalls durch andere Texte ersetzt werden. Dies ist allerdings nur das kleinste und einfachste Werkzeug aus dem großen Suchen/Ersetzen-Werkzeugkasten.

Reguläre Ausdrücke

InDesign bietet ab der Version CS3 neben der normalen Suchfunktion die Suche mit *Regulären Ausdrücken* bzw. GREP*. Hier haben Sie die Möglichkeit, nach Textmustern (z.B. Zahlen, Leerräumen, Großbuchstaben etc.) zu suchen. Ein Beispiel ist die Suche nach allen Zahlen von 0 bis 9 mit einem einzigen Suchbefehl. Die gefundenen Zeichen können dann bei der Ersetzung wiederverwendet werden. Stellen Sie sich vor, Sie müssten alle Seitenverweise in einem Dokument von der Form (S. 100) in die Form → Seite 100 bringen. Mit GREP brauchen Sie dafür nur eine Ersetzungsanweisung, weil die unterschiedlichen Seitenzahlen in der Ersetzung wieder eingesetzt werden.

Seit InDesign CS4 gibt es zusätzlich die Möglichkeit, in Absatzformaten so genannte GREP-Stile festzulegen. Hier werden bestimmte Zeichenfolgen definiert, die besonders formatiert werden sollen. Typisch wäre die Kleinerschaltung von Ziffern im Buchsatz oder die Hervorhebung von Namen. In InDesign CS6 wurden zwei Befehle hinzugefügt, die bei der Suche nach Leerräumen und Absätzen hilfreich sind.

Auf den ersten Blick sehen die Suchanfragen unübersichtlich und kompliziert aus. Dies ist der Abstraktion geschuldet, die gleichzeitig erst die Mächtigkeit der Suchausdrücke ermöglicht. Solange die Anfragen präzise und logisch formulierbar sind, kann man praktisch alles finden.

Reguläre Ausdrücke sind mit einigen Grundlagen schnell zu erlernen. Für den produktiven Einsatz ist es auch gar nicht notwendig, die letzten Details und Kniffe zu kennen. Statt theoretisch Möglichkeiten zu diskutieren, ist es oft hilfreicher, das grundsätzliche Prinzip zu verstehen und bei schwierigen Anfragen zu wissen, wie man weiterkommt.

* Die genaue Bedeutung von GREP ist umstritten, ich finde **General Regular Expression Parser** passend. Wer es genau wissen will, kann unter ↗ 122 nachsehen.

Im Folgenden stelle ich typische Suchanfragen vor, die man am besten mit GREP lösen kann. Im weiteren Verlauf des Buches werden noch komplexere Such- bzw. Ersetzungsanweisungen (→ Kapitel 10) und Möglichkeiten zum Skripten von GREP-Anfragen besprochen (→ Unterkapitel 4.8).

1.1 Der Suchen/Ersetzen-Dialog

Wenn Sie mit BEFEHL + F den Suchen/Ersetzen-Dialog aufrufen, sehen Sie die vier Registerkarten für die verschiedenen Suchtypen und die Bedienelemente für die Ausführung der Anweisungen. Der Dialog ist übrigens nicht modal, d. h., Sie können zwischen Dokument und Dialog hin und her wechseln und z. B. per Copy & Paste Text, Leerräume und Sonderzeichen aus dem Dokument übernehmen.

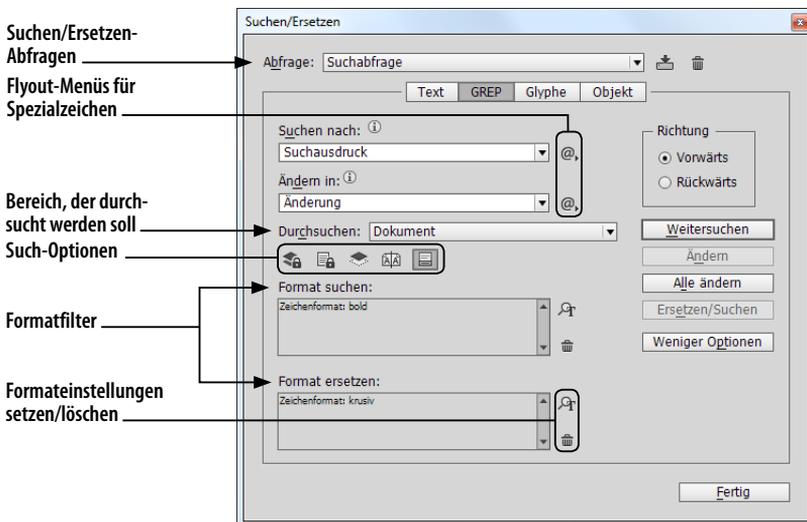


Abb. 1
Suchen/Ersetzen-Dialog

Mit den Registerkarten TEXT und GREP können Texte durchsucht und Ersetzungen vorgenommen werden. Der Aufbau und die Such-Optionen der beiden sind nahezu gleich. In der Registerkarte GREP hat man die Möglichkeit, mit regulären Ausdrücken zu arbeiten.

Mit der Registerkarte GLYPHE können bestimmte Zeichen anhand ihrer ID, mit der Registerkarte OBJEKT Rahmen anhand von Attributen gesucht werden. Diese Registerkarten werden in diesem Buch nicht vorgestellt, weil man sie in der Praxis nur selten benötigt.

Die Registerkarte GREP enthält die Texteingabefelder SUCHEN NACH und ÄNDERN IN, die Buttons zur Einstellung der Such-Optionen und die Felder für die Formateinstellungen.

Die Registerkarte GREP

Im Suchen- bzw. Ändern-Feld werden die regulären Ausdrücke für Suchanfragen bzw. Ersetzungsanweisungen eingetragen. Für die Arbeit

mit Regulären Ausdrücken werden einige Platzhalter und Zeichen mit einer speziellen Bedeutung benötigt. Diese finden Sie neben den Feldern im Flyout-Menü des @,-Buttons.

Unter den Texteingabefeldern kann mit einer Dropdown-Liste der Bereich, der durchsucht werden soll, festgelegt werden. Im Normalfall finden Sie hier die Auswahl zwischen allen Dokumenten und dem aktiven Dokument. Wenn sich die Einfügemarke beim Aufruf des Dialogs in einem Text befindet, finden Sie zusätzlich die Möglichkeiten, den aktuellen Textabschnitt oder den Textabschnitt von der Einfügemarke bis zum Ende zu durchsuchen. Der Textabschnitt beinhaltet den Text des Textrahmens oder den Text der miteinander verketteten Rahmen. Falls Sie einen Text ausgewählt haben, gibt es noch die Möglichkeit, die Auswahl zu durchsuchen.

Such-Optionen

Unterhalb der Dropdown-Liste können die Such-Optionen eingestellt werden. Mit ihnen kann festgelegt werden, welche Bereiche durchsucht werden sollen. Im Normalfall ist nur die Suche innerhalb von Fußnoten eingeschaltet. Weiterhin kann man für die Suche und Ersetzung auch die Musterseiten  und ausgeblendete Ebenen  mit einbeziehen. Gesperrte Ebenen bzw. Objekte  und gesperrte Textabschnitte  können nur bei der Suche berücksichtigt werden. Zu den Such-Optionen zählt auch die ab InDesign CC vorhandene Möglichkeit, VORWÄRTS bzw. RÜCKWÄRTS zu suchen.

Im Gegensatz zu der Textsuche sind in der Registerkarte GREP keine Buttons für die Suche nach einem ganzen Wort und für die Beachtung der Groß-/Kleinschreibung vorhanden. Diese Einstellungen werden bei Regulären Ausdrücken im Suchausdruck festgelegt.

Die Suchen/Ersetzen-Funktion von InDesign geht weit über das Suchen und Ersetzen von Buchstaben, Wörtern oder Satzteilen hinaus. InDesign bietet die Möglichkeit, die Suche mit der Angabe von Textformatierungen und Absatzattributen weiter zu verfeinern. Bei der Ersetzung können alle diese Formateinstellungen dann geändert bzw. gesetzt werden. Diese Einstellungen werden in den Feldern FORMAT SUCHEN bzw. FORMAT ERSETZEN angezeigt. Mit der Lupe  gelangen Sie zur Formatauswahl, mit dem Mülleimer  können Sie die Formateinschränkungen wieder entfernen.

Abb. 2
Fette Textstellen
kursiv formatieren



Vermutlich ist Ihnen auch schon aufgefallen, dass der Dialog alle Einstellungen nach dem Schließen beibehält. InDesign vergisst die Einstellungen erst beim Neustart des Programms.

Es können auch komplette Suchen/Ersetzen-Abfragen abgespeichert werden. Dazu finden Sie im Dialog eine Dropdown-Liste und den Button  zum Speichern bzw.  Löschen der aktuellen Abfrage. Interessante Abfragen kann man so für die spätere Verwendung speichern. Beim Skripting kann man auf die so gespeicherten Abfragen wieder zugreifen. Mit Hilfe des Skripts *ChainGREP.jsx* können diese Abfragen verkettet und gesammelt ausgeführt werden (→ Unterkapitel 10.1).

Suchen/Ersetzen-
Abfragen abspeichern

1.2 Die Suche mit Regulären Ausdrücken

Zum Einstieg empfiehlt es sich, einfach Suchanfragen auszuprobieren. Wenn Sie alle Vorkommen in einem Dokument ersetzen, sollten Sie allerdings Vorsicht walten lassen, da die Gefahr besteht, dass ungewollte Ersetzungen durchgeführt werden.

Zur besseren Übersichtlichkeit sind alle Suchanfragen und Ersetzungsanweisungen in diesem Kapitel grau hinterlegt.

Das Einfachste vorab: Ganz normale Suchtexte funktionieren auch mit GREP. Wenn nach `Max` gesucht wird, kann der Name durch `Moritz` ersetzt werden.

1.2.1 Zeichen mit spezieller Bedeutung

Mit Regulären Ausdrücken kann man variable Zeichenketten suchen und Suchergebnisse präzise eingrenzen. Diese Flexibilität wird dadurch erreicht, dass einige Zeichen eine besondere Bedeutung haben.

Wenn man z.B. im Texteingabefeld Suchen `~$` eingibt, findet InDesign geschützte Leerzeichen `^` und nicht die beiden Zeichen `~` und `$`. Hier wird ein Sonderzeichen durch die Kombination aus dem Tilde-Zeichen mit einem anderen Buchstaben dargestellt. Das Tilde-Zeichen ist hier ein *Metazeichen*. Metazeichen stehen nicht für sich selbst, sondern haben eine besondere Bedeutung.

Metazeichen

Neben dem Tilde-Zeichen gibt es noch einige weitere Zeichen, die eine spezielle Bedeutung haben. So findet man z.B. mit dem Punkt `.` nicht nur das Satzzeichen, sondern fast jedes beliebige Zeichen. Wenn nur nach einem Punkt gesucht wird, muss die besondere Bedeutung aufgehoben werden. Dies geschieht mit einem vorangestellten Backslash. Nach einem Punkt muss also mit `\.` gesucht werden. Diese Aufhebung der Bedeutung von Zeichen mit spezieller Bedeutung nennt man *Maskierung* (engl. escape) – sie wird übrigens ganz ähnlich auch beim Skripting benötigt.

Maskierung

In Regulären Ausdrücken gibt es Zeichen, um Sonderzeichen zu erzeugen, die Suche genauer zu steuern oder etwas im Suchausdruck zu markieren. Diese Zeichen müssen alle mit dem Backslash maskiert werden, wenn sie buchstäblich gefunden werden sollen. Der Backslash zählt auch dazu und wird mit `\\` gefunden.

! Metazeichen bei der Suche mit GREP

! In InDesign CS6 funktioniert \\$ nicht. Verwenden Sie [\$]

\ . * + ? () { } [^ \$ | ~

Wenn man wo bist du? finden möchte, muss man nach wo bist du\? suchen. Zunächst ist es nur wichtig, die Metazeichen zu kennen. Sie werden im Verlauf des Kapitels vorgestellt.

Die Idee, die Suchanfragen mit Metazeichen zu verfeinern, gab es auch schon vor InDesign CS3 in der normalen Textsuche. Dort gibt es z.B. die Möglichkeit, mit ^p nach einem Absatzende oder mit ^? nach einem beliebigen Zeichen zu suchen. Hier wird das Caret-Zeichen (Zirkumflex) ^ zur Kennzeichnung von Zeichen mit spezieller Bedeutung verwendet. Wenn man es finden möchte, muss man es ebenfalls mit ^^ maskieren. Diese Möglichkeiten sind für die Textsuche in der Registerkarte TEXT erhalten geblieben.

Die Problematik mit den Metazeichen ist der einzige Grund, zuweilen noch die normale Suche in der Registerkarte TEXT zu verwenden. Wenn man nur eine bestimmte Zeichenfolge sucht, muss man sich dort (mit Ausnahme des Caret-Zeichens) nicht um die Maskierung kümmern.

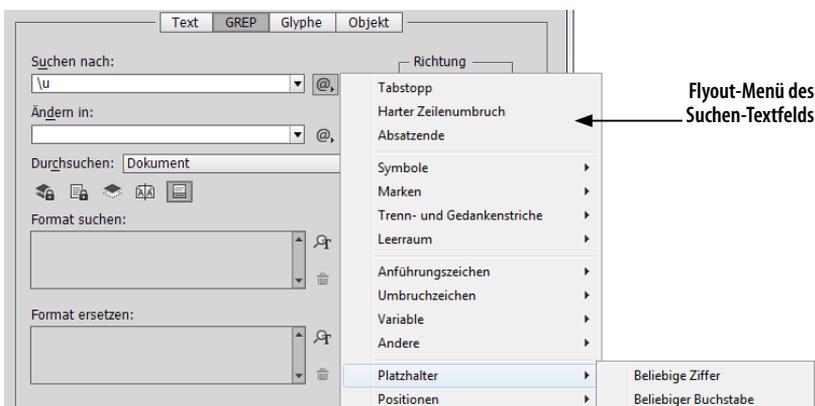
1.2.2 Variable Zeichen auswählen

Platzhalter

Die wichtigsten GREP-Befehle sind die so genannten *Platzhalter* (engl. wildcard), mit denen unterschiedliche Zeichen gefunden werden. Sie ermöglichen eine unscharfe Suche nach Zeichengruppen wie allen Kleinbuchstaben oder allen Ziffern.

Die im Folgenden vorgestellten Zeichen mit einer speziellen Bedeutung finden Sie im Flyout-Menü @, neben dem Suchen/Ersetzen-Textfeld.

Abb. 3
Platzhalter
Flyout-Menü



Beliebige Zeichen
finden

Der wichtigste und einfachste Platzhalter ist der Punkt `.`. Mit ihm kann jedes beliebige Zeichen, mit Ausnahme vom Absatzende (hierzu zählen der normale Zeilenumbruch ¶ und auch alle anderen Umbruchzeichen wie z. B. Seitenumbruch) und dem harten Zeilenumbruch (Soft-Return) ↵, gefunden werden. Mit `..` würde man zwei beliebige aufeinanderfolgende Zeichen finden. Platzhalter repräsentieren eine bestimmte Menge an Zeichen und werden deswegen als *Zeichenklassen* bezeichnet.

Die Suche nach `Max.` findet nicht nur `Max.`, sondern auch `Maxi` mit `i` und ohne Punkt sowie alle anderen Formen von `Max0` bis `Max:`. Mit der Suche nach `d.` könnte man z. B. alle Wörter finden, die mit `d` beginnen und mindestens drei Buchstaben lang sind.

Ein Platzhalter trifft genau ein Zeichen. Wenn Sie einen Punkt in das Suchen-Feld eingeben und die Suche mehrmals ausführen, wird ein Zeichen nach dem anderen gefunden. Um die Verwendung von Zeichenklassen zu verdeutlichen, greife ich auf das nächste Unterkapitel vor und führe die Möglichkeit ein, mit dem Plus-Zeichen mehrere Zeichen vom gleichen Typ zu suchen. Wenn man ein `+` hinter eine Zeichenklasse setzt, wird das erste und alle folgenden zusammenhängenden Zeichen der Zeichenklasse gefunden. Wenn man das Plus-Zeichen mit der Zeichenklasse Punkt kombiniert `.+`, werden entsprechend alle Zeichen bis zum Absatzende oder einem harten Zeilenumbruch gefunden.

Zeichenklassen

Neben dem Punkt gibt es noch einige andere nützliche Zeichenklassen. Diese werden alle aus Kombinationen mit dem Backslash gebildet.

Mit `\u` findet man einen beliebigen Großbuchstaben, mit `\l` einen Kleinbuchstaben. Mit `M\l+` findet man entsprechend alle Wörter, die mit einem großen `M` beginnen und denen Kleinbuchstaben folgen: `Max`, `Moritz`, `Montag`, `Mars` usw., aber nicht `MacOS`.

Entsprechend findet man mit `\u\l+` alle Wörter, die mit einem Großbuchstaben beginnen, mit `\l+` alle kleingeschriebenen.

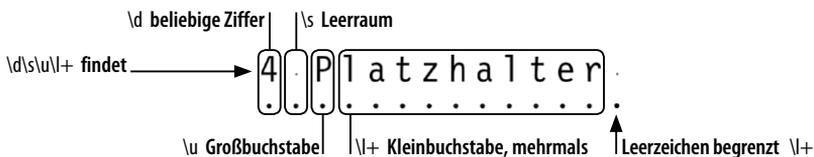


Abb. 4
Verhalten von
Platzhaltern

Eine weitere wichtige Zeichenklasse sind Zahlen. Alle Ziffern von 0 bis 9 können mit `\d` gefunden werden. Alle beliebigen Ziffern in einem Text finden Sie mit `\d+`. Später zeige ich noch die Möglichkeit, auch Ziffern mit Nachkommastellen in den Suchausdruck miteinzubeziehen.

Mit `\s` findet man alle Leerräume, also u. a. Leerzeichen, geschützte Leerzeichen `^`, Festabstände wie z. B. Halbgeviert `␣`, den Tabulator `»` und Zeilenschaltungen `¶` bzw. `↵`.

Mit InDesign CS6 hat Adobe die Platzhalter `\h` für alle horizontalen Leerräume (Leerzeichen, Tabulator etc.) und `\v` für alle vertikalen Leerräume (Absatzende und harter Zeilenumbruch) eingeführt. Bei anspruchsvollen Layouts, in denen mikrotypografische Anpassungen vorgenommen werden sollen, kann man mit `\h` sehr gut verschiedene Kombinationen suchen und durch eine Ersetzung vereinheitlichen. Mit dem Suchausdruck `\d\hmm` findet man z. B. alle Kombinationen einer

Die Platzhalter `\h` und `\v` wurden in InDesign CS6 eingeführt.

Zahl und der angehangenen Maßeinheit mm unabhängig vom Leer-
raum, der dazwischen verwendet wurde.

Zeichenklassen
negieren

Außerdem hat man die Möglichkeit, die Bedeutung der Zeichenklassen
umzudrehen, d.h. also alle Zeichen, die ihr *nicht* entsprechen, zu su-
chen. So kann man mit `\D` alles außer Ziffern finden – hier findet man
neben Buchstaben auch Leerräume, Satzzeichen etc.

Tab. 1
Oft verwendete
vordefinierte
Zeichenklassen

Zeichen- klasse	Beschreibung
.	Ein beliebiges Zeichen, aber nicht das Absatzende (Return) oder einen erzwungenen Zeilenumbruch (Soft-Return)
\u	Alle Großbuchstaben (engl. uppercase). Findet keine elektronisch versal TT gestellten Texte.
\l	Alle Kleinbuchstaben (engl. lowercase)
\d	Alle Ziffern (engl. digits) 0–9. Keine Brüche und nur elektronisch hoch- oder tiefgestellte Zahlen.
\s	Alle Leerräume, also Leerzeichen, alle Festabstände, Tabulatoren und Umbruchzeichen (engl. space). Fälschlicherweise auch das selten ver- wendete »Einzug bis hierhin«- Zeichen †.
\h	Alle horizontalen Leerräume: Leerzeichen, alle Festabstände und den normalen Tabulator (nicht den Tabulator für rechte Ausrichtung).

Zeichenauswahl

Im Flyout-Menü finden
Sie die Zeichenauswahl
unter **ENTSPRECHUNG** →
ZEICHENSATZ

Etwas genauer als mit Zeichenklassen kann man mit einer *Zeichen-
auswahl* suchen. Hier können die Zeichen, die gesucht werden sollen,
selbst festgelegt werden – man kann sozusagen seine eigene Zeichen-
klasse definieren. Eine Zeichenauswahl wird innerhalb von eckigen
Klammern `[]` festgelegt. So findet man mit dem Ausdruck `S[kc]ript` die
Variationen Skript und Script. Innerhalb der eckigen Klammern kön-
nen beliebig viele Zeichen eingetragen werden, die im Suchtreffer an
der entsprechenden Position vorkommen können. Ein anderes Beispiel
wäre die Suche nach `Buch[st][st]abe`, mit der man z.B. Buchstaben-
dreher zwischen s und t auflösen kann. Mit der Suche nach `[gjpqy]`
wiederum würde man alle Buchstaben mit Unterlängen finden – bei
kursiven Schriften kämen noch *f* und *ß* hinzu.

Noch flexibler wird eine Zeichenauswahl durch die Möglichkeit, Be-
reiche festzulegen. Bereiche werden mit einem Bindestrich zwischen dem
ersten und letzten Zeichen des Bereichs definiert. Die Anfrage `200[5-9]`
findet alle Jahreszahlen zwischen 2005 und 2009. Mit der Anfrage `[A-N]`
findet man alle Großbuchstaben zwischen A und N. Solche Bereiche kann
man gut zur Formatierung von Listen, in denen z.B. Produktnummern
nach einem bestimmten Schema festgelegt sind, verwenden. Vorstellbar
ist z.B. eine Produktnummer, die mit 1 oder 2 beginnt, dann mit einem
Buchstaben zwischen A und N weitergeht, gefolgt von noch zwei weite-
ren Ziffern. Der passende Ausdruck dafür wäre `[12][A-N]\d\d`.

Weiterhin können auch negative Zeichenmengen angegeben werden. Hier kann man Zeichen festlegen, die nicht vorkommen dürfen. Dies wird mit dem Zirkumflex bzw. Caret-Zeichen `^` erreicht. Die Anfrage `InDesign CS[^3-9]` findet alle InDesign CS-Versionen vor CS3, wobei die Version CS auch ohne die Ziffer 1 gefunden wird, da das folgende Zeichen – vermutlich ein Leerzeichen – ebenfalls nicht in der Menge aller Zeichen zwischen 3 und 9 enthalten ist. Genauso könnte man mit `Abbildung\h[^d]` alle Abbildungslegenden finden, die noch keine Nummerierung haben.

Innerhalb der Zeichenauswahl gibt es andere Metazeichen, hier haben nur die schließende eckige Klammer `]`, der Zirkumflex `^`, das Tilde-Zeichen `~` und der Bindestrich (Divis) `-` eine besondere Bedeutung. Für die Maskierung braucht man wieder den Backslash. Zwischen den eckigen Klammern sind nur die folgenden Zeichen reserviert:

`] ^ - \ ~`

Die im vorhergehenden Abschnitt genannten Metazeichen haben hier keine besondere Bedeutung mehr und müssen nicht maskiert werden. Die Suche nach `[.\~]` findet Punkte und Tilde-Zeichen. Die Maskierung des Bindestrichs kann am Ende der Zeichenauswahl entfallen `[.\~-]`.

Innerhalb einer Zeichenauswahl können auch Zeichenklassen eingesetzt werden. Mit `[\d,]` findet man Zahlen mit Nachkommastellen – eventuell muss noch der Punkt hinzugefügt werden.

Mit `[\dA-N]+` würde man wieder die oben erwähnte Produktnummer finden. Diese Anfrage ist aber weit weniger präzise, weil Reihenfolge und Länge der Nummer nicht festgelegt sind.

Mit dem Ausdruck `[\u1\~]+` könnte man die Zeichenklassen von Groß- und Kleinbuchstaben sowie den Bindestrich suchen, also praktisch alle Wörter. Vielleicht kennen oder finden Sie die Zeichenklasse `\w`, die »offiziell« für diesen Zweck gedacht ist. Mit `\w` findet man jedoch alle Buchstaben, Zahlen und den Unterstrich, für die Suche nach Wörtern ist das eher unbrauchbar.

Als weitere praktische Anwendung stelle ich die Suche nach einer E-Mail-Adresse vor. Dreh- und Angelpunkt ist das `@`-Zeichen. Vor dem `@` dürfen sich Buchstaben, Zahlen und Unterstriche befinden, die mit `\w` gefunden werden können. Da aber auch Punkte und Bindestriche erlaubt sind, muss eine eigene Zeichenklasse gebildet werden `[\w.-]+`. Nach dem `@`-Zeichen sind die gleichen Zeichen erlaubt. Zur genaueren Eingrenzung sollte die Domainendung, die sich hinter einem Punkt befindet und nur noch aus den Zeichen `a-z` bestehen darf, angefügt werden `\.[a-z]+`. Zusammengefügt ergibt sich der Ausdruck `[\w.-]+@[\w.-]+\.[a-z]+`. Bitte beachten Sie, dass E-Mail-Adressen sehr komplex sein können und mit der hier vorgestellten GREP-Suche nur die üblichen Varianten gefunden werden.

Negative
Zeichenauswahl

 **Metazeichen bei
der Festlegung einer
Zeichenauswahl**

Wörter finden

E-Mail-Adressen finden

Alternativen

Wenn man nicht nach Zeichenmengen, sondern nach variierenden Wortteilen suchen möchte, müssen die Variationen in runde Klammern geschrieben und mit dem senkrechten Strich | unterteilt werden.

Die Suche nach `(Java|Apple|VB)Script` findet alle drei Programmiersprachen innerhalb des Textes. Genauso kann natürlich auch einfach nach zwei alternativen Wörtern gesucht werden – `(Nela Malou|Silke)` findet beide Vorkommen im Text. Es liegt auf der Hand, dass diese Suchanfrage deutlich präziser ist als die ebenfalls erfolgreiche Suche nach allen Buchstaben der beiden Namen mit `[NMSe|aouik]+`.

1.2.3 Sonderzeichen

Viele Zeichen können nicht direkt in die Such- bzw. Ersetzungsanfrage eingegeben werden. So ist es z.B. nicht möglich, eine Zeilenschaltung direkt in das Texteingabefeld einzugeben. Bei anderen Zeichen ist dies zwar prinzipiell möglich, aber ein geschütztes Leerzeichen wäre im Texteingabefeld nicht von einem normalen Leerzeichen zu unterscheiden. Die dritte Klasse bilden die Zeichen mit einer besonderen Bedeutung. Wenn die eigentlichen Zeichen gefunden werden sollen, müssen diese, wie am Anfang beschrieben, mit dem Backslash maskiert werden.

Tabulator und Zeilenumbrüche werden in Regulären Ausdrücken mit dem Backslash erzeugt. Einen oder mehrere Tabulatoren findet man z.B. mit dem Suchausdruck `\t+`.

Tab. 2
Standardsonderzeichen
für die Suche mit
Regulären Ausdrücken

GREP	Zeichen	Beschreibung
<code>\t</code>	»	Tabulator
<code>\r</code>	¶, ↵, ⏎ ...	Absatzende bzw. Zeilenschaltung (Return), aber auch Rahmenumbruch, Seitenumbruch etc.
<code>\n</code>	↵	Erzwungener Zeilenwechsel (Soft-Return)

Alle weiteren Sonderzeichen werden mit der Tilde ~ erzeugt. Alle Sonderzeichen, die so erzeugt werden, finden Sie nach Bereichen sortiert im zweiten und dritten Bereich des Flyout-Menüs neben dem Suchenfeld.

Abb. 5
Sonderzeichen-
Auswahl im Suchen/
Ersetzen-Dialog



Nach einem geschützten Leerzeichen sucht man z. B. mit `~S`, nach bedingten Trennzeichen mit `~-`. Interessant ist die Möglichkeit, dass auch Indexmarken `~I`, verankerte Objekte `~a` und Textvariablen `~v` gesucht werden können.

Grundsätzlich können Sie alle Zeichen auch per Copy & Paste in die Texteingabefelder des Suchen/Ersetzen-Dialogs einfügen. InDesign erzeugt dann automatisch die richtige Kombination mit der Tilde. Die folgende Tabelle zeigt einige oft gebräuchliche Sonderzeichen, eine vollständige Übersicht finden Sie im Anhang A1.

GREP	Zeichen	Beschreibung
<code>~S</code>	^	Geschütztes Leerzeichen
<code>~m</code>	□	Geviert
<code>~></code>	◌◌◌◌	Halbgeviert
<code>~<</code>	◌◌◌◌	Achtelgeviert
<code>~b</code>	¶	Standardzeilenumbruch
<code>~=</code>	—	Halbgeviertstrich, Gedankenstrich
<code>~-</code>	-	Bedingter Trennstrich
<code>~y</code>	↵	Tabulator rechte Ausrichtung
<code>~I</code>	⋈	Indexmarke
<code>~a</code>	⌘	Marker für verankerte Objekte
<code>~v</code>		Alle Textvariablen
<code>\~</code>	~	Tilde

Tab. 3
InDesign-
Sonderzeichen in
Regulären Ausdrücken

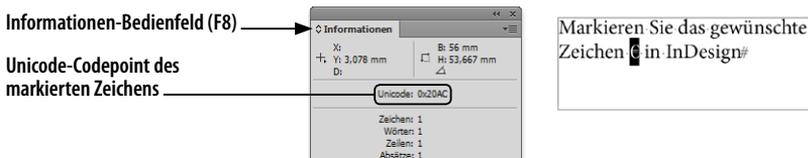
Kombinationen mit der Tilde gelten nur in InDesign und können nicht in anderen Programmen, die Reguläre Ausdrücke unterstützen, eingesetzt werden.

Sonderzeichen und Zeichenklassen können in einem alternativen Suchausdruck oder einer eigenen Zeichenklasse kombiniert werden. So könnte man der Zeichenklasse `\h` mit dem Ausdruck `(\h|~y)` bzw. `[\h~y]` den fehlenden Tabulator rechte Ausrichtung hinzufügen.

Neben diesen festgelegten Zeichen gibt es auch die Möglichkeit, ein Zeichen über den *Unicode*-Codepoint festzulegen. Dazu muss der hexadezimale Codepoint des Zeichens bekannt sein. Nach dem Zeichen `¾` suchen Sie mit `\x{00BE}`, das €-Symbol finden Sie mit `\x{20AC}`. Den Codepoint eines markierten Zeichens kann man in InDesign in der Mitte des Informationsfensters ablesen (FENSTER → INFORMATIONEN). Weitere Informationen zu Unicode finden Sie auf Seite 148. Im Kapitel 10.7 werden weitere Suchtechniken vorgestellt.

++
Unicode-Zeichen
finden

Abb. 6
Informationen-
Bedienfeld



Wenn der Codepoint dort zu sehen ist, können Sie das Zeichen aber genauso gut per Copy & Paste in die Texteingabefelder des Dialogs befördern. Die Definition eines Zeichens mit `\x{####}` benötigt man nur, wenn ein Unicode-Zeichen gesucht wird, dessen Codepoint aus einer Tabelle oder einem anderen Programm bekannt ist.

1.2.4 Wiederholungen

Eine weitere wichtige Technik in Regulären Ausdrücken ist es, die Anzahl von Wiederholungen eines Zeichens oder einer Zeichenklasse festzulegen. Das Wiederholungszeichen `+`, mit dem man mehrere Zeichen finden kann, wurde bereits verwendet.

Abb. 7
Wiederholung
Flyout-Menü

Wiederholung	Null oder ein Mal _____?
Entsprechung	Null oder mehrere Male _____*
Modifizierer	Ein oder mehrere Male _____+
Posix	Null oder ein Mal (kürzeste Entsprechung) _____??
	Null oder mehrere Male (kürzeste Entsprechung) _____*?
	Ein oder mehrere Male (kürzeste Entsprechung) _____+?

Wiederholungszeichen

Immer wenn man nicht genau weiß, ob ein Suchbegriff oder Zeichen im Text vorkommt, kann man das Wiederholungszeichen `?` verwenden (im Menü NULL ODER EIN MAL). So könnte man z.B. alle Vorkommen von `das` und `dass` mit der Suchanfrage `dass?` prüfen. Mit `(Max)? Mustermann` finden Sie jede Erwähnung von Herrn Mustermann – mit und ohne Vorname. Das Wiederholungszeichen, auch *Quantifizierer* genannt, bezieht sich immer auf das vorhergehende Zeichen, eine Zeichenklasse oder das vorhergehende Wort in runden Klammern. Es kann nicht alleine stehen.

Mit einem `+` wird festgelegt, dass ein Suchausdruck mindestens einmal vorkommen muss, aber beliebig oft vorkommen darf (im Menü EIN ODER MEHRERE MALE). Wenn Sie z.B. mehrere Leerzeichen hintereinander zu einem einzigen vereinheitlichen wollen, kann man mit `+` eine beliebige Anzahl von Leerzeichen finden. Verwenden Sie dazu nicht `\s+`, da in der Zeichenklasse auch die Absatzschaltungen enthalten sind und Sie eventuell die Kombination Leerzeichen Return entfernen würden.

Ein Plus bietet sich oft an, um den Rest eines Absatzes oder den variablen Text zwischen zwei bekannten Zeichenfolgen zu finden. Wenn ein beliebiger Text, z.B. die Beschreibung nach einer Produktnummer

oder vor dem Preis, gesucht wird, helfen die speziellen Zeichenklassen meist nicht weiter, da die Beschreibung fast alle Zeichen enthalten kann. Angenommen einer wie oben geformten Produktnummer folgt eine Produktbeschreibung, z. B. 1G78 Staubsauger. Die Nummer inklusive Beschreibung finden Sie jetzt mit `[12][A-N]\d\d\h.+` – es wird sogar 1G79 Staubsauger mit Beutel gefunden oder 1H32 Wischmops.

Wenn man nicht genau weiß, ob das Zeichen oder die Zeichenklasse überhaupt vorkommt, kann mit einem `*` auch nach beliebig vielen Vorkommen gesucht werden (im Menü NULL ODER MEHRERE MALE, korrekt wäre die Bezeichnung BELIEBIG OFT). Der Suchbegriff kann keinmal, einmal oder beliebig oft vorkommen. Im Unterschied zum `?` findet man mit dem `*` eine beliebige Anzahl von Zeichen – mit `jaa*` findet man ja, jaa, jaaa usw. Dies ist z. B. nötig, um nach einer Zahl zu suchen, die eventuell Nachkommastellen haben kann. Der Suchausdruck `\d+,\?\d*` findet sowohl 15 als auch 14,9999, wobei die Anzahl der Nachkommastellen nicht relevant ist. Im Suchausdruck könnte man auch das Komma mit einem `*` versehen. Dies wäre aber nicht besonders präzise, weil bei einer Zahl mit Nachkommastellen genau ein, bei einer Zahl ohne Nachkommastellen genau kein Komma enthalten ist.

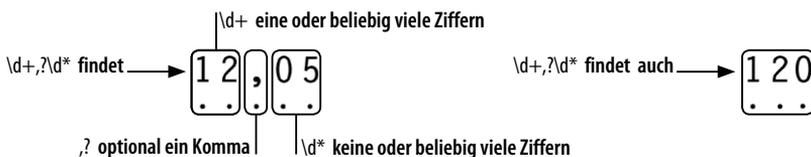


Abb. 8
Wiederholungen

Achten Sie darauf, dass das Sternchen in Kombination mit nur einem Zeichen immer trifft – eine solche Suche ist also nicht sinnvoll. Beispielsweise findet `x*` an jeder Position des Textes entweder ein `x` oder kein `x`.

Wiederholungszeichen	Beschreibung
<code>?</code>	Das Zeichen oder die Zeichenklasse ist optional, d. h., sie kann einmal oder keinmal vorkommen.
<code>+</code>	Der Suchbegriff muss mindestens einmal, kann aber beliebig oft vorkommen.
<code>*</code>	Der Suchbegriff kann beliebig oft vorkommen – also auch keinmal.

Tab. 4
Wiederholungsfaktoren
in Regulären
Ausdrücken

Suchverhalten der Wiederholungszeichen

Die Wiederholungszeichen `+` und `*` finden immer so viele Zeichen wie möglich. Sie suchen so lange weiter, bis sie den größtmöglichen passenden Suchausdruck gefunden haben. Das ist auch der Grund, warum man mit `.\+` nicht den ersten Satz eines Absatzes auswählen kann, wie man vielleicht vermuten könnte. In diesem Ausdruck sammelt das `+` beliebige Zeichen vom Absatzbeginn bis zum letzten Vorkommen eines

InDesign versucht normalerweise den längsten Treffer zu ermitteln.

Das Fragezeichen hat je nach Kontext unterschiedliche Bedeutungen.

Punkts und damit meist den ganzen Absatz. Dieses Verhalten wird als *gierig* (engl. greedy) bezeichnet.

Um die Wiederholungszeichen zu zügeln, kann ein Fragezeichen verwendet werden. Hinter einem Wiederholungszeichen bedeutet es, dass nur die kleinstmögliche Anzahl Treffer gefunden werden soll. Mit dem Ausdruck `.+?\.` wird nur noch der erste Satz eines Absatzes gefunden. Das Fragezeichen hat also mehrere Bedeutungen: Hinter einem Zeichen oder einer Zeichenklasse bewirkt es, dass das Suchzeichen im Treffer ein- oder keinmal vorkommen darf. Hinter einem Wiederholungszeichen regelt es, dass so wenige Zeichen wie möglich gefunden werden.

Das Fragezeichen sollte immer eingesetzt werden, wenn die Gefahr besteht, dass die nächste Fundstelle nach einem Wiederholungszeichen nicht eindeutig ist. Wenn z.B. in Klammern gesetzter Text gesucht wird (also dieser hier), ist es durchaus möglich, dass innerhalb des Absatzes noch ein weiterer Text in runden Klammern steht (hier ist der zweite). Wenn man nun mit einem gierigen Wiederholungszeichen `\(.+\)` sucht, findet man den gesamten Text von der ersten bis zur letzten runden Klammer. Hier sollte man stattdessen mit dem Ausdruck `\(.+?\)` arbeiten. Achten Sie auch darauf, dass die spezielle Bedeutung der Klammern durch die Maskierung mit dem Backslash aufgehoben werden muss.

Genau Anzahl der Vorkommen festlegen

Noch genauer kann mit geschweiften Klammern `{}` festgelegt werden, wie oft ein Zeichen oder eine Zeichenklasse vorkommen darf. Eine Zahl innerhalb der geschweiften Klammern legt fest, wie viele Vorkommen des Zeichens oder der Zeichenklasse erlaubt sind. So findet man mit `\d{4}` alle vierstelligen Zahlen.

Wenn man einen Wiederholungsbereich festlegen möchte, muss die minimale und maximale Anzahl der Vorkommen angegeben werden. Beide Werte werden durch ein Komma getrennt. Zwei- oder dreistellige Zahlen findet man entsprechend mit `\d{2,3}`.

Tab. 5
Genau Anzahl von Wiederholungen festlegen

Wiederholungen	Beschreibung
<code>{n}</code>	Der Ausdruck muss exakt n-mal vorkommen.
<code>{n,m}</code>	Der Ausdruck muss mindestens n-mal und darf maximal m-mal vorkommen.
<code>{n,}</code>	Der Ausdruck muss mindestens n-mal und darf öfter vorkommen.

1.2.5 Genau Positionen ermitteln

Suchbereich

Eine Suche kann oft durch Positionsangaben im Suchbereich noch genauer eingegrenzt werden. Standardmäßig endet der Suchbereich nach dem Absatzende oder dem harten Zeilenumbruch – dem ersten Zeichen, das nicht von `\` gefunden wird.