Lokator: Regular Expression

Der Lokator Regular Expression findet reguläre Ausdrücke in den OCR-Textzeilen des Dokumentes.

Das Ergebnis dieses Lokators ist das gefundene Suchmuster.

> Wie reguläre Ausdrücke funktionieren, ist nicht Bestandteil dieser Dokumentation, dafür gibt es im Internet sehr viele gute Beispiele und Möglichkeiten zum Testen von regulären Ausdrücken, z.B. https://regex101.com/

> Wichtig: Es können beliebig viele reguläre Ausdrücke konfiguriert werden, diese werden dann automatisch **oder** verknüpft gesucht.

> Wichtig: reguläre Ausdrücke werden in Squeeze case insensitive gesucht, das bedeutet, Groß-Klein Schreibung muss nicht extra berücksichtigt werden.

> Wichtig: Die regulären Ausdrücke werden in der Reihenfolge in der diese angelegt sind gesucht, das bedeutet für einen regulären Ausdruck der bereits gefunden wurde, kann kein weiterer, in der Liste nachfolgender regulärer Ausdruck gefunden werden.

Klassische Beispiele für den Einsatz von Lokatoren für reguläre Ausdrücke:

> Wichtig: für einige dieser regulären Ausdrücke müssen Ersetzungen konfiguriert werden

Beispiel	Wert	regulärer Ausdruck	Leerzeichen ignorieren
IBAN	DExx xxxx xxxx xxxx xxxx xx	(DE\d{20})	ja
Ust-ID	DE xxxxxxxx	((DE)([1-9]\d{8}))	ja
Beträge	100,00 oder 1.000,00	([-\+]?[0-9]{1,3}([ ]?[,\.]?[ ]?[0-9]{3})*[ ]?[,\.][ ]?[0-9]{2}[-\+]?(?![0-9.,]))	nein
Datum	01.01.2020	([0-9]{1,2}\.[0-9]{1,2}\.[0-9]{4})|([0-9]{1,2}\.[0-9]{1,2}\.[0-9]{2})	ja

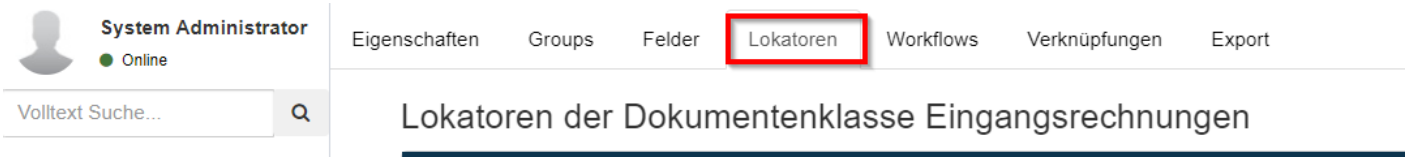
Beispiel	Wert	regulärer Ausdruck	Leerzeichen ignorieren
Namen	Achim Redmann	(Achim Redmann)	nein
Telefonnr.	+4940359840001	([+]?[0-9]{8,15})	ja
Emailadresse	info@dexpro-solutions.de	([a-zA-Z0-9_\-.]{2,30}@[a-zA-Z0-9-]{2,30}\.[a-zA-Z]{2,3})	nein
URL	www.dexpro-solutions.de	(www\.[a-zA-Z0-9-]{2,30}\.[a-zA-Z]{2,3})	nein

beispielhafte Einrichtung eines neuen Lokators zur Erkennung einer Bestellnummer anhand eines regulären Ausdrucks

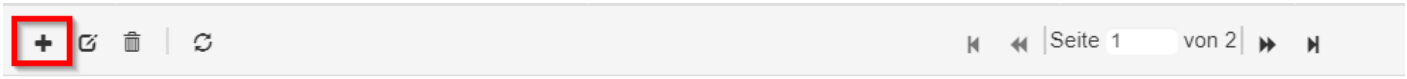
Im folgenden Beispiel soll eine auf dem Dokument befindliche 10-stellige numerische Bestellnummer erkannt werden.

Dazu bietet sich der Lokator für reguläre Ausdrücke sehr gut an.

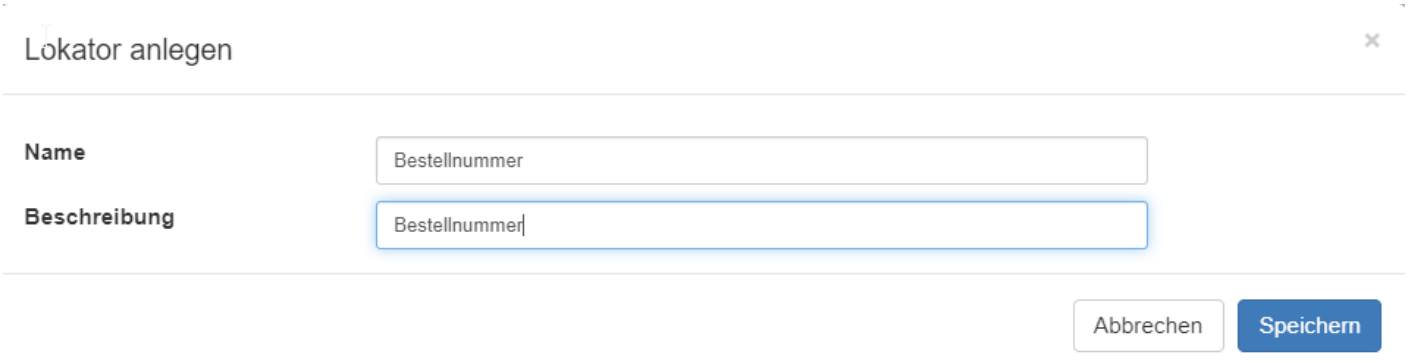
Im Konfigurationsdialog einer Dokumentenklasse auf den Reiter Lokatoren klicken.



In der unteren Bildschirmleiste das + Symbol Klicken um einen neuen Lokator anzulegen.



Im sich daraufhin öffnenden Fenster den technischen Namen und den Anzeige-Namen des neuen Lokators angeben und auf den Button Speichern klicken.

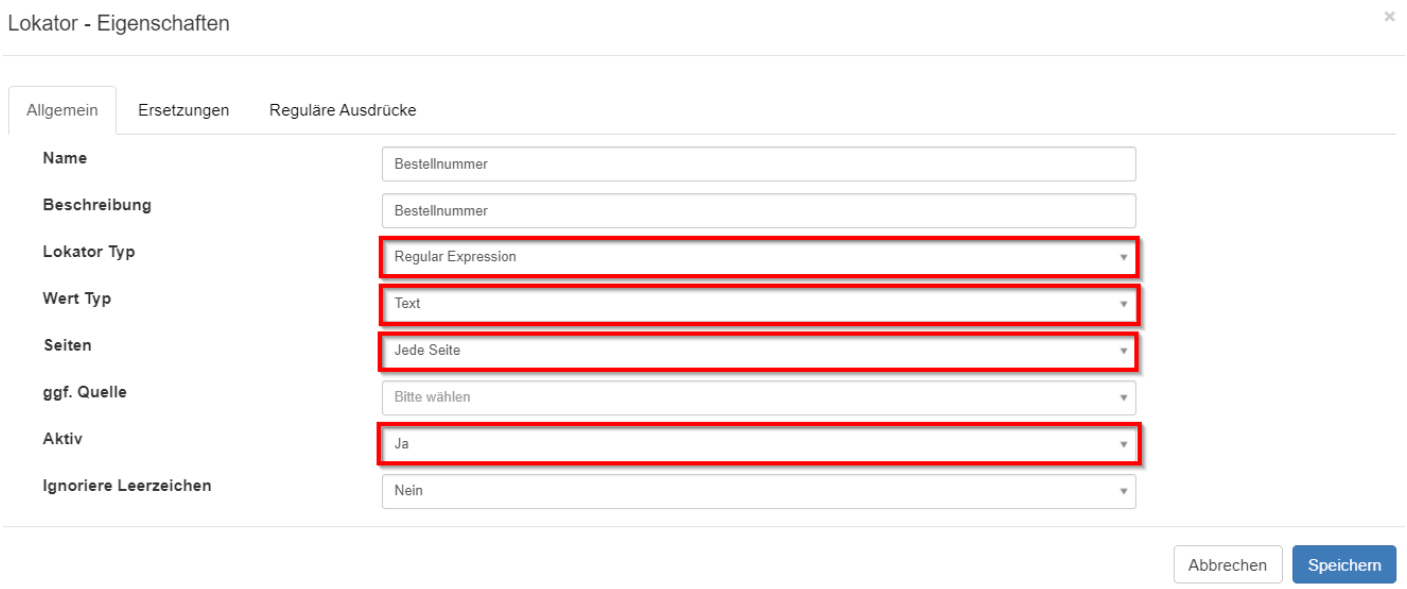


Der neue Lokator erscheint nun in der Liste der Lokatoren. Auf diesen Eintrag Doppelklicken.

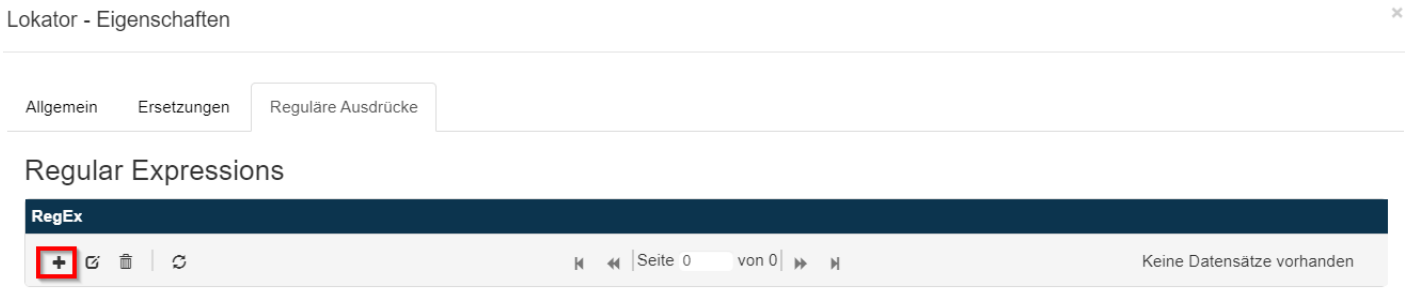
Lokatoren der Dokumentenklasse Eingangsrechnungen

ID	Name	Beschreibung	Typ
21	HeaderTotalAmount	Überschrift Gesamtpreis	KeyWord to Value
22	LineItems	Positionen	Search for line items
23	Bestellnummer	Bestellnummer	Regular Expression

Im sich daraufhin öffnenden Dialog den Lokator auf Aktiv setzen. Der Lokator-Typ Regular Expression ist bereits vorausgewählt. Die Erkennung der Bestellnummer soll auf jeder Seite durchgeführt werden und der Wert Typ den wir erkennen wollen ist Text.

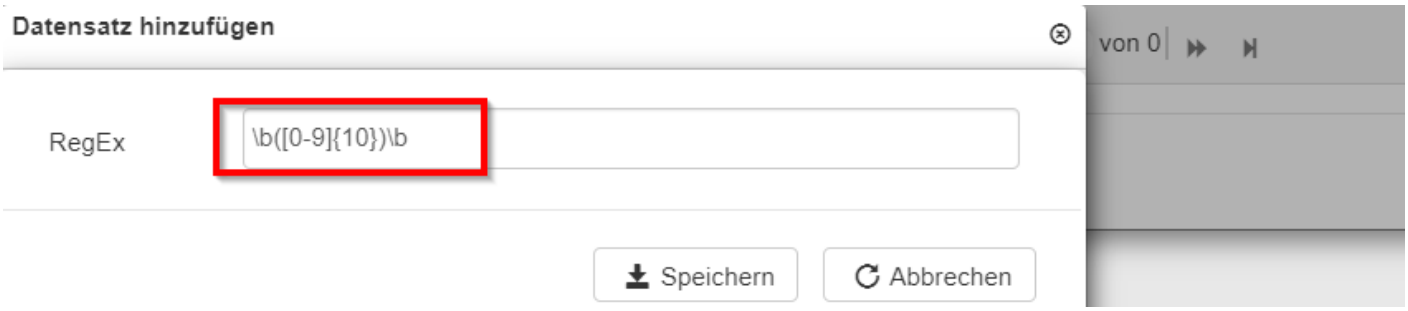


Dann klicken wir auf den Reiter Reguläre Ausdrücke um den Reg-Ex für den Lokator zu konfigurieren. Hier wieder das + Symbol klicken um einen neuen regulären Ausdruck zu konfigurieren.



Der reguläre Ausdruck für eine freistehende 10 stellige Nummer könnte folgend konfiguriert werden: \b([0-9]{10})\b

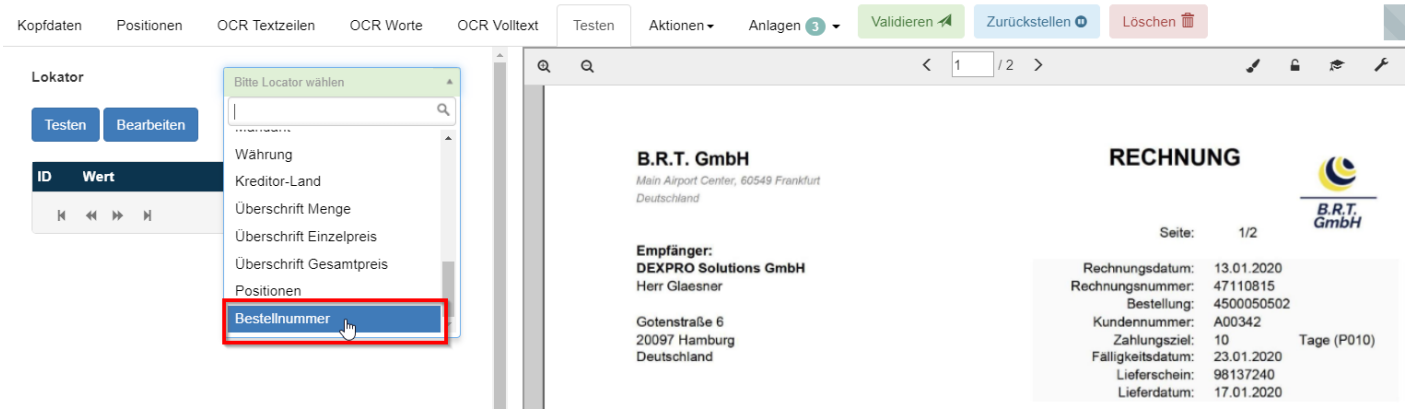
dann auf Speichern klicken.



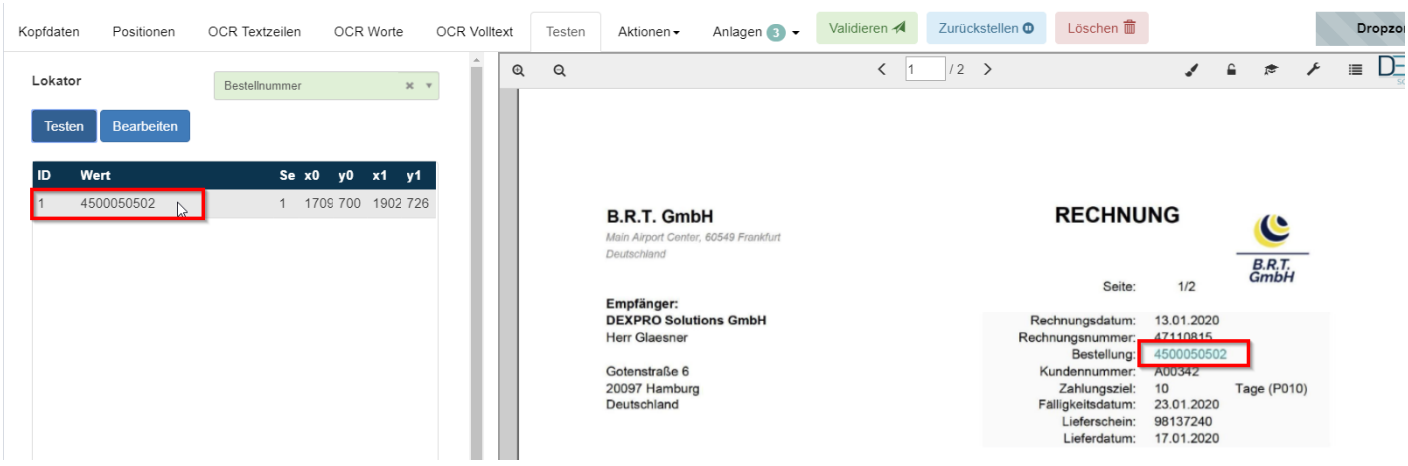
Der neue Reg-Ex steht jetzt in der Liste der regulären Ausdrücke.



Nachdem der neue Lokator konfiguriert wurde, kann man den Lokator im Testmodus testen. Dazu wird das Dokument mit der Bestellnummer geöffnet und auf den Reiter Testen geklickt. Der neue Lokator kann jetzt in der Liste im Auswahlfenster ausgewählt werden.



Mit einem Klick auf den Testen Button wird der neue Lokator getestet. Das Ergebnis der Lokatorsuche wird wieder links in der Liste dargestellt, die Fundstelle wird auf dem Dokument markiert.



Jetzt müsste der Benutzer den neuen Lokator noch dem jeweiligen Dokumenten-Feld für die Bestellnummer zuweisen. Dies ist im Kapitel Dokumentenfelder beschrieben.

Wie die erkannte Bestellnummer gegen entsprechende Stammdaten plausibilisiert werden kann, ist in der Beschreibung für den Lokator Search for DB Link data beschrieben.

Revision #13
Created 2020-03-06 14:48:19 UTC by Achim Redmann
Updated 2020-03-09 14:58:40 UTC by Achim Redmann