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
Namen	Max Mustermann	(Max Mustermann)	nein

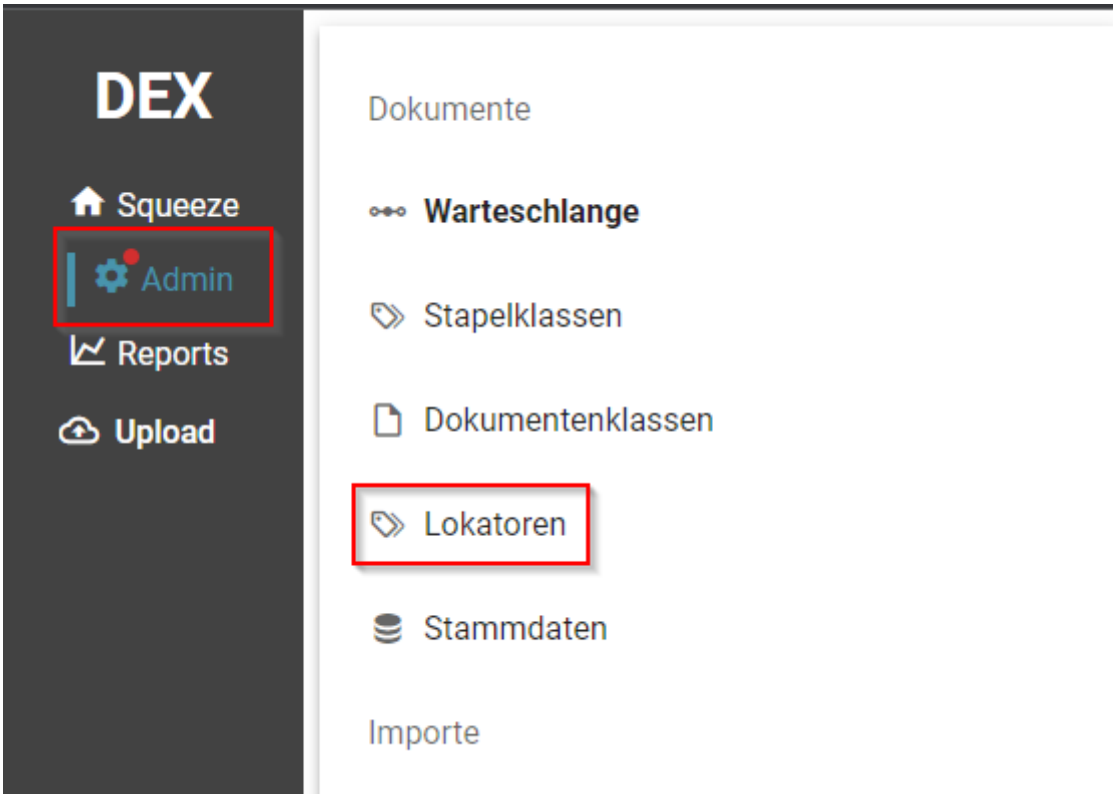
Beispiel	Wert	regulärer Ausdruck	Leerzeichen ignorieren
Telefonnr.	+4940359840001	([+]?[0-9]{8,15})	ja
Emailadresse	info@dexpro-solutions.de	([a-zA-Z0-9_\-.]{2,30}@[a-zA-Z0-9-]{2,30}\.[a-zA-Z]{2,3})	nein
URL	www.dexpro-solutions.de	(www\.[a-zA-Z0-9-]{2,30}\.[a-zA-Z]{2,3})	nein

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

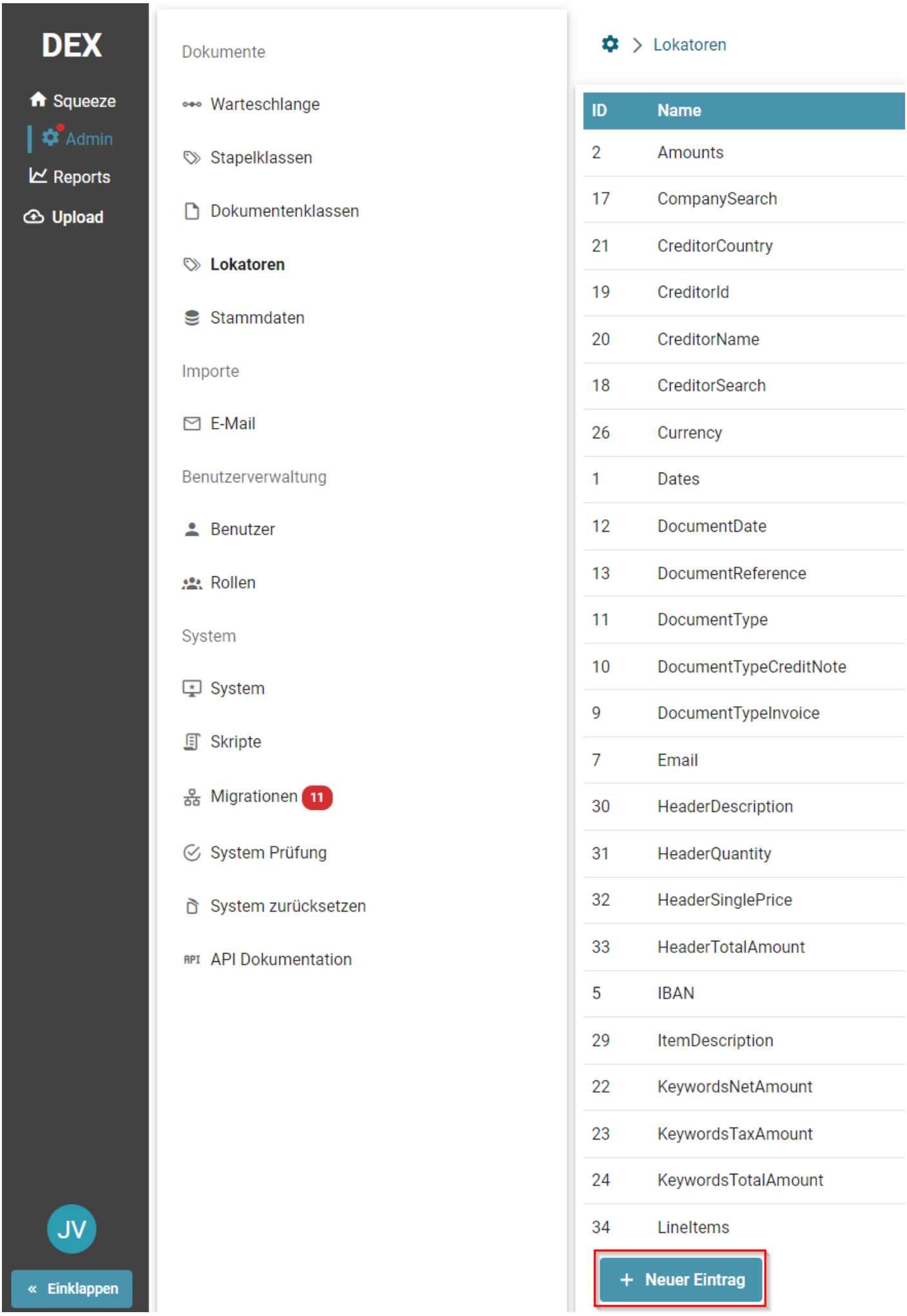
Im folgenden Beispiel soll eine auf dem Dokument befindliche 10-stellige numerische Bestellnummer erkannt werden.

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

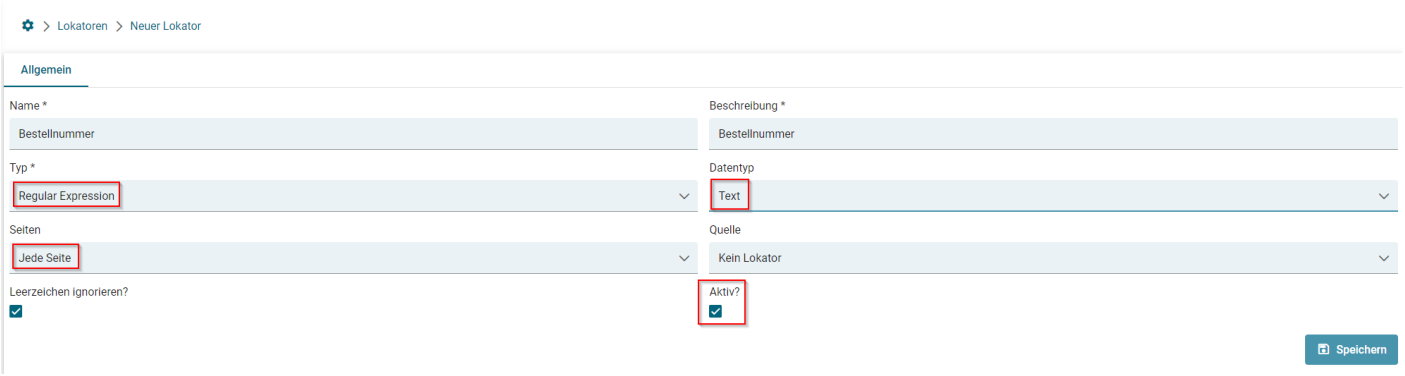
Um ins Lokatoren-Menü zu gelangen Klicken Sie auf Admin und danach auf Lokatoren.



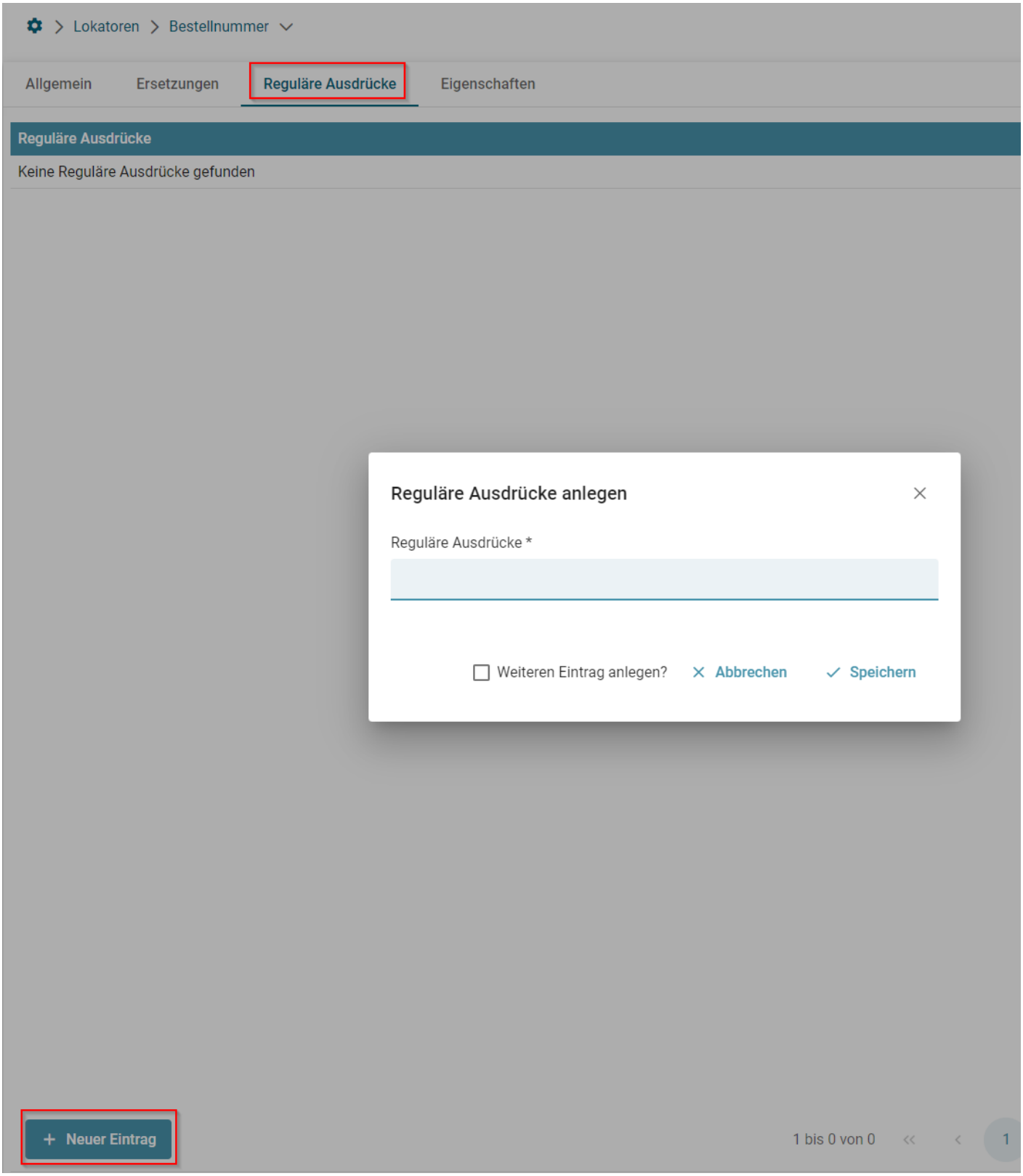
In der unteren Bildschirmleiste das Symbol "Neuer Eintrag" klicken um einen neuen Lokator anzulegen.


DEX
Squeeze
Admin
Reports
Upload
JV
« Einklappen
Dokumente
Warteschlange
Stapelklassen
Dokumentenklassen
Lokatoren
Stammdaten
Importe
E-Mail
Benutzerverwaltung
Benutzer
Rollen
System
System
Skripte
Migrationen 11
System Prüfung
System zurücksetzen
API Dokumentation
Lokatoren
ID
Name
2 Amounts
17 CompanySearch
21 CreditorCountry
19 CreditorId
20 CreditorName
18 CreditorSearch
26 Currency
1 Dates
12 DocumentDate
13 DocumentReference
11 DocumentType
10 DocumentTypeCreditNote
9 DocumentTypeInvoice
7 Email
30 HeaderDescription
31 HeaderQuantity
32 HeaderSinglePrice
33 HeaderTotalAmount
5 IBAN
29 ItemDescription
22 KeywordsNetAmount
23 KeywordsTaxAmount
24 KeywordsTotalAmount
34 LineItems
+ Neuer Eintrag

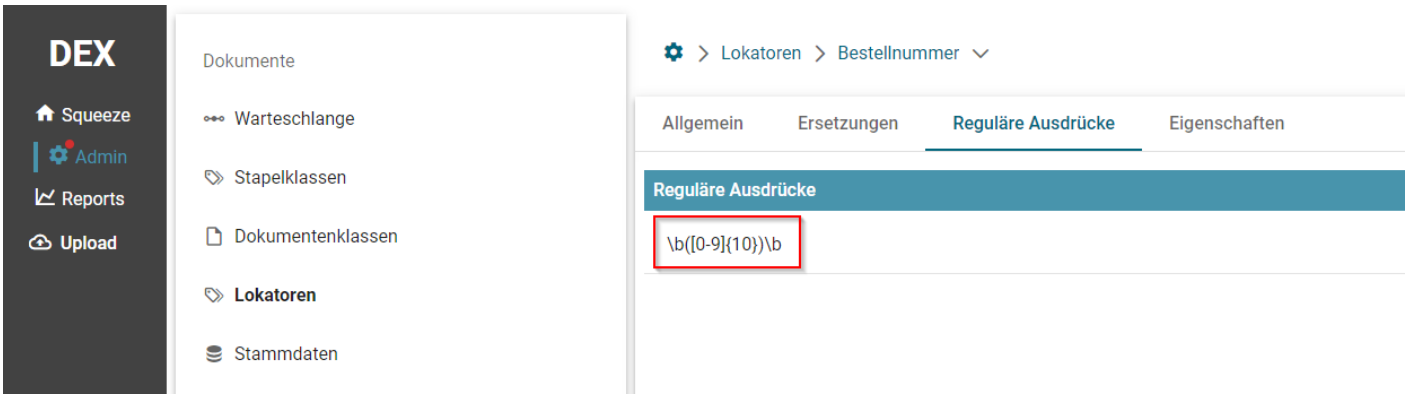
Im sich daraufhin öffnenden Dialog den technischen Namen, den Anzeige-Namen des neuen Lokators angeben und den Lokator auf Aktiv setzen. Der Lokator-Typ Regular Expression ist bereits vorausgewählt. Die Erkennung der Bestellnummer soll auf jeder Seite durchgeführt werden und der Wert Typ den wir erkennen wollen ist Text. Danach wird der Lokator gespeichert.



In der danach erscheinenden Ansicht klicken wir auf den Reiter Reguläre Ausdrücke um den Reg-Ex für den Lokator zu konfigurieren. Hier wieder das Symbol "Neuer Eintrag" 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.



Wie dieser Lokator getestet werden kann finden Sie auf der Seite "Testen von Lokatoren".

Revision #4
Created 2022-03-29 14:05:00 UTC by Jan Vogl
Updated 2023-02-07 11:19:30 UTC by Jan Vogl