

Unterschiede zu Squeeze 1

Diese Seite beschreibt relevante Unterschiede zwischen der Extraktion mittels Lokatoren zwischen Squeeze 1 und Squeeze 2.

Globale Lokatoren

In Squeeze 1 wurden Lokatoren immer als Teil einer Dokumentenklasse konfiguriert. Das hatte zur Folge, dass Lokatoren nur in ihrer Dokumentenklasse genutzt werden konnten. Wenn ein System über mehrere Dokumentenklassen verfügt hat, so mussten übliche Lokatoren (Email / IBAN usw.) in jeder Dokumentenklasse erneut angelegt werden.

In Squeeze 2 hingegen werden Lokatoren unabhängig von der Dokumentenklasse verwaltet. Somit ist es möglich, dass bspw. nur ein einziger IBAN-Lokator zu konfigurieren ist und dieser in beliebig vielen Dokumentenklasse wiederverwendet werden kann.

Diese Änderung gilt im Prinzip für alle Lokatoren. Der Tabellen-Lokator (Search for line items) hingegen ist der einzige Lokator, bei welchem diese Umstellung aktuell (Q3 2022) nur teilweise zutrifft.

Besonderheit Tabellen-Lokator (Search for line items)

Zusammenfassung / TLDR:

- Der Tabellen-Lokator (Search for line items) ist aktuell **nicht** global.
- Für jede zu extrahierende Tabelle muss ein eigener Tabellen-Lokator konfiguriert werden.
- Überschriften- und Wert-Lokatoren von Tabellen können mehrmals verwendet werden.

Kontext

Wie in [Globale Lokatoren](#) beschrieben sind Lokatoren global, d. h. sie können mehrfach in unterschiedlichen Dokumentenklassen für diverse Felder wiederverwendet werden. Bei der Extraktion von Tabellen mittels des Tabellen-Lokators (Search for line items) allerdings trifft diese Umstellung nur teilweise zu.

Falls Sie nur wissen möchten, wie die Positionserkennung zu konfigurieren ist, dann lesen Sie bitte die [Dokumentation zum entsprechenden Lokator](#). Dort haben wir dokumentiert, wie in Squeeze 2 das System zu konfigurieren ist.

Andernfalls wird im Folgenden erläutert, wie sich die Konfiguration zwischen Squeeze 1 und 2 unterscheidet, damit Administratoren die Umstellung einfacher fällt.

Funktionsweise des Lokators

Zuerst muss die Funktionsweise des Lokators (egal ob Squeeze 1 oder 2) verstanden werden. Vereinfacht arbeitet der Tabellen-Lokator in folgenden Schritten:

- Spalten der Tabelle über ihre Überschriften finden (*nutzt Überschriften-Lokator*)
- Werte innerhalb der gefundenen Spalten finden (*nutzt Wert-Lokator*)
- Gefundene Werte zu Zeilen zusammenfassen

Wichtig ist hierbei zu verstehen, dass der Lokator also auf Basis von Überschriften- und Werte-Lokatoren arbeitet. Aus diesem Grund, muss der Tabellen-Lokator wissen, für welche Tabelle dieser ausgeführt wird, um dann wiederum über die Tabellen-Spalten der Tabelle zu erfahren, welche Überschriften- und Werte-Lokatoren zu nutzen sind.

Der Tabellen-Lokator ist nicht global

Die beschriebene Funktionsweise des Tabellen-Lokators setzt voraus, dass bei diesem konfiguriert werden muss, für welche Tabelle dieser ausgeführt wird.

Während das aus funktionaler Sicht kein Problem darstellt, ist der Lokator allerdings somit der einzige Lokator, welcher de facto **nicht** global einsetzbar ist, was bei der Konfiguration zu Irritation führen kann:

Während alle anderen Lokatoren wiederverwendet werden können, muss für die Extraktion einer Tabelle je ein eigener Tabellen-Lokator genutzt werden.

Roadmap bzgl. der Positionserkennung

Abschließend sei erwähnt, dass die Besonderheit des Positionslokators bekannt ist und das Konzept auch zukünftig überarbeitet werden soll. Ziel dabei soll ist die Einfachheit und Nachvollziehbarkeit der Konfiguration.

Änderungen an dem System werden entsprechend in der Dokumentation (insbesondere [Lokator: Search for DB linked data](#)) fortlaufend festgehalten.

Revision #4

Created 31 August 2022 15:32:39 by Fabian Terstegen

Updated 7 February 2023 11:19:30 by Fabian Terstegen