

# Stapelklassen

Konfigurations-Anleitungen im Kontext von Stapelklassen.

- [Stapelklassen](#)
- [Stapelklassen anlegen](#)
- [Stapelklassen-Eigenschaften](#)
- [Stapelklasseneigenschaften OCR-Texterkennung](#)

# Stapelklassen

Stapelklassen erfüllen in Squeeze mehrere Funktionen.

## **1. Stapelklassen sind Gruppierungen von Dokumentenklassen**

In einer Stapelklasse können mehrere Dokumentenklassen eingerichtet werden und für die Klassifizierung zusammengefasst werden. Z.B bei der Einrichtung eines Mailrooms.

## **2. Stapelklassen für Barcode-Erkennung**

Die Barcode-Erkennung wird zum Trennen nach Barcode oder zur Erkennung von Anhängen nach Barcode verwendet. Die Definition der Barcodes und ihrer Eigenschaften finden Sie auf der Seite Stapelklassen-Eigenschaften.

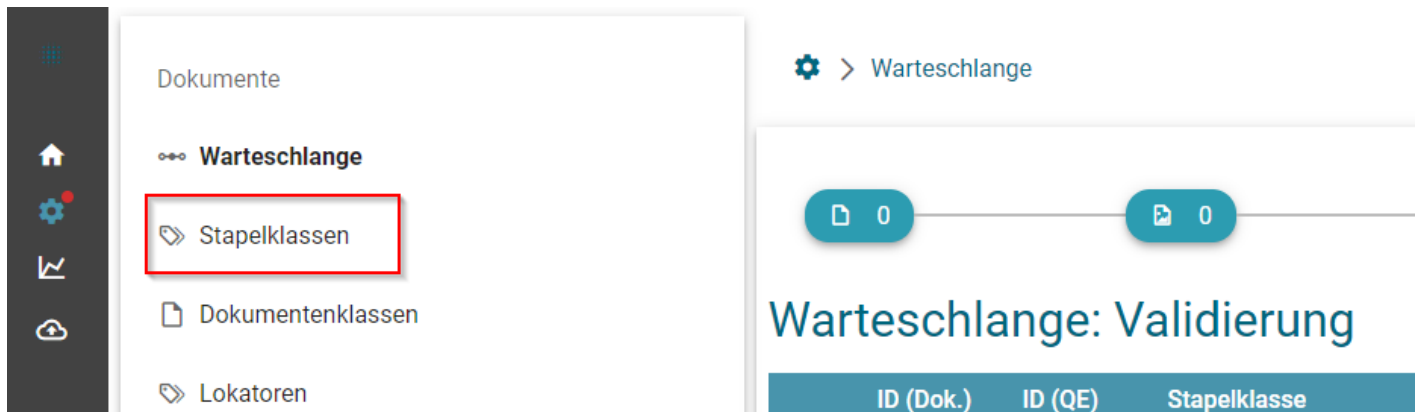
## **3. Stapelklasse für OCR-Sprache**

Im Standard werden Deutsch und Englisch verwendet. Dies kann erweitert werden, da es in manchen Sprachen bestimmte Sonderzeichen bzw. andere Buchstabensätze gibt. Die Definition der OCR-Sprachen und ihrer Eigenschaften finden Sie auf der Seite Stapelklassen-Eigenschaften.

# Stapelklassen anlegen

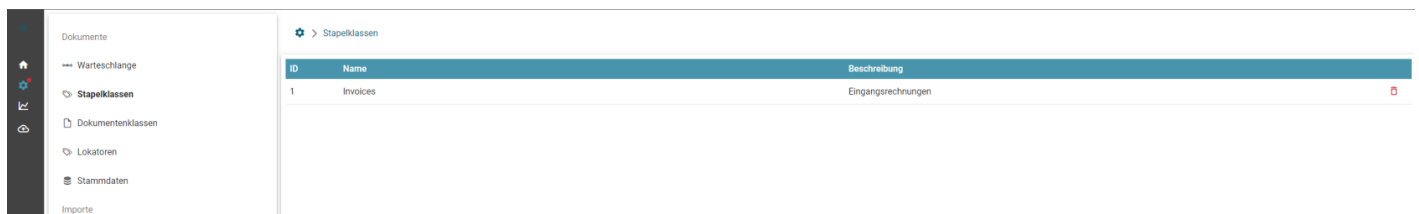
Nach der Anmeldung in Squeeze können Benutzer mit administrativen Berechtigungen die Konfiguration für Stapelklassen aufrufen.

## Liste der Stapelklassen



Nach dem Klick auf den Reiter Stapelklassen öffnet sich die Stapelklassenübersicht. Hier werden alle aktuell konfigurierten Stapelklassen angezeigt. Die Stapelklasse Invoice wird zusammen mit dem Invoice Template ausgeliefert.

In der Liste der Stapelklassen kann man die Konfiguration mit einem **Klick** auf den Eintrag öffnen.



## Neue Stapelklasse anlegen

Neue Stapelklassen können mithilfe des Symbols "Neuer Eintrag" angelegt werden. Daraufhin öffnet sich ein Dialog in dem der technische Namen und der Anzeigenamen eingegeben werden kann.



Ab hier können folgende Konfigurationen durchgeführt werden:

1. OCR
2. Klassifikationen
3. Eigenschaften

## OCR

Eine Stapelklasse enthält eine OCR-Konfigurationsoberfläche, welche spezifische Eigenschaften enthält, die konfiguriert werden können. Die vorhandenen Einstellungen der OCR werden ebenfalls unter den Eigenschaften einer Stapelklasse aufgelistet.

The screenshot shows a web application interface for configuring OCR settings. On the left is a dark sidebar with a navigation menu containing icons and labels for various system functions: Dokumente, Warteschlange, Stapelklassen, Dokumentenklassen, Lokatoren, Stammdaten, Importe, E-Mail, Benutzerverwaltung, Benutzer, Rollen, System, System, Export & Import Konfigurationen, Skripte, Jobverwaltung, Migrationen, System Prüfung, and System zurücksetzen. The main content area has a breadcrumb trail: > Stapelklassen > Eingangsrechnungen. Below this is a tabbed interface with four tabs: Allgemein, OCR (selected), Klassifikationen, and Eigenschaften. The OCR tab contains several configuration fields: OCREngine\* (dropdown menu showing 'default'), OCRForce (dropdown menu showing 'Ja'), PSM-Mode (dropdown menu showing '3'), OCRPageLimit (dropdown menu showing '1-3'), OCRLanguage (dropdown menu), PDFConversion (dropdown menu), and PDFProcessor (dropdown menu showing 'PDFBox'). Each dropdown menu has a small 'X' icon and a downward arrow. At the bottom right of the configuration area is a button labeled 'Speichern' (Save).

## Klassifikationen

Mit einem Klick auf "Neuer Eintrag" können der Stapelklasse Dokumentenklassen zugeordnet werden.

Dokumente

- Warteschlange
- Stapelklassen**
- Dokumentenklassen
- Lokatoren
- Stammdaten

Importe

- E-Mail

Benutzerverwaltung

- Benutzer
- Rollen

System

- System
- Export & Import Konfigurationen
- Skripte
- Jobverwaltung
- Migrationen
- System Prüfung
- System zurücksetzen

> Stapelklassen > Eingangsrechnungen

AllgemeinOCR**Klassifikationen**Eigenschaften

ID	Dokumentenklasse
1	Eingangsrechnungen

+ Neuer EintragKlassifizierung testen

## Eigenschaften

Über das Symbol " Neuer Eintrag" können der Stapelklasse spezifische Eigenschaften hinzugefügt werden.

Dokumente

Warteschlange

Stapelklassen

Dokumentenklassen

Lokatoren

Stammdaten

Importe

E-Mail

Benutzerverwaltung

Benutzer

Rollen

System

System

Export & Import Konfigurationen

Skripte

Jobverwaltung

Migrationen

System Prüfung

System zurücksetzen

TA

Stapelklassen > Eingangsrechnungen

Allgemein

OCR

Klassifikationen

Eigenschaften

Eigenschaft	Wert		
PSM-Mode	3		
SpaceMaxWidth	55		
OCRForce	true		

+ Neuer Eintrag

1 bis 3 von 3 << < 1 > >> 25

## Eigenschaft speichern



Eigenschaft \*



BarcodeEngine

BarcodeImageColor

EasAllowDocumentWithoutAttachment

ExportAfterExtraction

# Stapelklassen-Eigenschaften

Mit Stapelklasseneigenschaften lassen sich allgemeine Konfigurationen pflegen, welche sich auf die Verarbeitung der Dokumente auswirken, die zu der jeweiligen Stapelklasseneigenschaft gehören.

## Eigenschaften

### AttachmentBarcodePattern und SplitBarcodePattern

RegEx zum Erkennen von Barcodenummern für Folgendes Beispiel:



Hier wäre ein RegEx zum Erkennen von 10 Nummern nötig. `([0-9]{10})`

Folgen die Nummern jedoch einem bestimmten Muster kann man die Suche eingrenzen. Z.B. `0000012345`, `0000012346`, `00000123457`, etc... `(00000[0-9]{5})`

### AttachmentBarcodeType

Unter AttachmentBarcodeType versteht man die möglichen Barcode-Typen, die verwendet werden können um Anhänge zu erkennen. Z.B. : Typ `CODE_128`

Die unterstützten Barcode-Typen werden in der UI angezeigt. Für ältere Systeme finden Sie weiter unten eine Liste.

### BarcodeEngine

#### Informationen zur Datenverarbeitung mit Barcodes

ZXING	Wird als Standard verwendet. Liefert aktuell die Besten Ergebnisse.
ZBAR	Kann als Alternative benutzt werden, falls ZXING kein Ergebnis liefert.

SOFTEK	Engine, die ab Squeeze 2.4 verfügbar ist und eine separate gültiger Lizenz erfordert.
ALL	Es werden beide Engines nacheinander verwendet (führt zu längerer Laufzeit). Verfügbar ab Version 2.0

## BarcodeImageColor

original	Das Originalbild wird unverändert verwendet.
grayscale	Das Bild wird in Graustufen konvertiert.
black	Das Bild wird in Schwarz/Weiß konvertiert.

## EasAllowDocumentWithoutAttachment

Im Regelfall wird die Archivierung nur zugelassen, wenn am Vorgang auch ein Dokument vorhanden ist. Sollten auch Dokumente archiviert werden können muss diese Eigenschaft auf `true` gesetzt werden. Standard ist hier der Wert `false`.

## ExportAfterExtraction

Hier kann festgelegt werden ob Dokumente in die Validierung gelangen oder direkt exportiert werden sollen. Für den direkten Export wird `true` eingetragen. Soll das nicht passieren tragen Sie `false` ein. Standard ist hier der Wert `false`.

## FilterDuplicateEmailAttachments

Beim Import wird für jedes Vorgangsdokument ein Hashwert erstellt und geprüft.

Sollte in einer neuen Email das exakt gleiche Dokument erneut importiert werden, wird geprüft ob es diesen Hash bereits gibt. Wenn ja und die Stapelklasseneigenschaft gesetzt ist wird dieses Dokument nicht importiert.

Wert `true` gilt als Standard für aktiviert und `false` für Deaktiviert.

## FilterDuplicateEmails

Beim Import wird für jede EML-Datei ein Hashwert erstellt und geprüft.

Sollte eine neue exakt gleiche Email erneut importiert werden, wird geprüft ob es diesen Hash bereits gibt. Wenn ja und die Stapelklasseneigenschaft ist gesetzt, wird diese Email nicht importiert.

Wert `true` gilt als Standard für aktiviert und `false` für Deaktiviert.

## IgnoreMandatoryFieldCheckForExport



Wenn der Automatische Export konfiguriert worden ist ( `ExportAfterExtraction` ) und diese Stapelklasseneigenschaft ist gesetzt, dann wird beim automatischen Export die Pflichtfeldprüfung ignoriert. Im Standard ist das Ignorieren der Felder deaktiviert. (Wert `false`). Um die Pflichtfeldprüfung beim automatischen Export zu deaktivieren ist diese Eigenschaft auf `true` zu setzen.

## OCR

Einstellungen bzgl. der OCR / Texterkennung werden hier näher dokumentiert:

[OCR-Stapelklasseneigenschaften in Squeeze](#)

## SoapAllowIncomingDocumentWithoutAttachment

Ist diese Eigenschaft auf "`true`" gesetzt können über den Soap-Server Dokumente angelegt werden ohne das ein Bild übergeben werden muss.

## SoapIncomingFileTypeFilter

Mit dieser Eigenschaft kann entschieden werden welche Dateitypen (PDF,TIF,JPEG,etc...) zugelassen werden. Ist diese Eigenschaft nicht gesetzt, werden nur PDF-Dateien zugelassen. Die Trennung erfolgt durch Semikolon. Die Dateieindungen sind Case-Insensitiv.

## SpaceMaxWidth

Bis zu welcher Anzahl an Pixeln soll ein Leerzeichen erkannt werden. Ein Beispiel im Standard wären 55 Pixel. Als Wert ist hier die Anzahl an Pixeln einzutragen.

## SplitBarcodeType

Unter SplitBarcodeType versteht man die möglichen Barcode-Typen, die verwendet werden können um ein neues Dokument zu erkennen. Z.B. : Typ `CODE_128`

Die unterstützten Barcode-Typen werden in der UI angezeigt. Für ältere Systeme finden Sie weiter unten eine Liste.

## SplitFixPages

Hierbei wird eine fixe Seitentrennung nach der angegebenen Seite konfiguriert. Z.B.: Trenne Jedes Dokument nach der zweiten (2) Seite.

## SkipXmlExtraction

Diese Stapelklasseneigenschaft kann mit `true` oder `false` bzw. mit `ja` oder `nein` konfiguriert werden. Diese Eigenschaft steuert das Überspringen der Werte-Extraktion von XML-Information. Ist diese Eigenschaft mit `true` belegt so überspringt der Extraktions-Schritt die Verarbeitung der XML-Information. Ist diese Eigenschaft mit `false` belegt, wird die Verarbeitung wie gewohnt

durchgeführt. Ist die Eigenschaft nicht konfiguriert so geht der Extraktions-Schritt davon aus, dass alle XML-Informationen nach Möglichkeit ausgewertet werden.

Diese Eigenschaft hat einen Effekt wenn das Dokument ein reines XML-Dokument ist. Sollte die Eigenschaft auf "ja"/"false" stehen, so werden reine XML-Dokumente wie XRechnung mit einer entsprechenden Fehlermeldung im 1. Verarbeitungsschritt abgelegt.

## XmlValidationReport

Diese Einstellung steuert ob im ersten Schritt der Dokumentenverarbeitung(Initialisierung-Schritt) ein KoSIT-Validierungs-Report erzeugt werden soll und kann mit `true` oder `false` bzw. mit `ja` oder `nein` konfiguriert werden.

Erfordert die Server/Mandanten-Konfiguration des Digivoice-Dienstes

## AI-Extraction

Mit dieser Stapelklasseneigenschaft kann die KI gestützte Extraktion und OCR aktiviert werden. Diese Stapelklasseneigenschaft kann mit `true` oder `false` bzw. mit `ja` oder `nein` konfiguriert werden.

Zusätzlich zu dieser Eigenschaft muss noch das AI-Extraction-Model angegeben werden, mit der die Extraktion und OCR durchgeführt werden soll.

Diese Eigenschaft benötigt gültige Zugangsdaten zur Autorisierung. Diese können bei der DEXPRO beauftragt werden.

## AI-Extraction-Model

Mit dieser Stapelklasseneigenschaft kann das Model der KI gestützten Extraktion und OCR ausgewählt werden.

Zur Auswahl stehen:

`invoice`

Dieses Modell ist speziell für Eingangsrechnungen trainiert. Dabei kann ein vom Standard abweichendes Mapping in dem UserExit `BeforeAiMapping` angegeben werden. Dieses Modell überspringt den OCR Schritt und liefert all seine Ergebnisse im Extraktionsschritt. Die OCR wird jedoch nur übersprungen, wenn die Dokumentenklasse bereits feststeht. Andernfalls wird erst die standardmäßig hinterlegte OCR genutzt um den Vorgang als Eingangsrechnung zu klassifizieren.

`legacy`

Dieses Modell kann mehrere Dokumententypen auslesen. Dazu gehören neben Eingangsrechnungen auch Liefer- und Bestellscheine.

Diese Eigenschaft benötigt gültige Zugangsdaten zur Autorisierung. Diese können bei der DEXPRO beauftragt werden.

## AsyncExportAfterValidation

Wenn diese Eigenschaft aktiviert wird, werden Dokumente nach der manuellen Validierung im Hintergrund exportiert.

Dieses Feature ist nützlich bei Export-Schnittstellen, die sehr langsam sind und erlaubt es Validierern das nächste Dokument zu validieren, während im Hintergrund das vorher validierte exportiert wird.

Fehler, die während des Exports auftreten, sorgen dafür, dass das Dokument erneut im Validierungs-Schritt angezeigt wird, damit eine manuelle Fehlerbehandlung / Support möglich ist.

# Barcodetypen

Der Wert für den Barcode Typ ist case-sensitiv (Groß-Kleinschreibung beachten!).

- AZTEC
- CODABAR
- CODE\_39
- CODE\_93
- CODE\_128
- COMPOSITE
- DATABAR
- DATA\_MATRIX
- DATABAR\_EXP
- EAN\_2
- EAN\_5
- EAN\_8
- EAN\_13
- ITF
- ISBN\_10
- ISBN\_13
- MAXICODE
- PDF\_417
- QR\_CODE
- RSS\_14

- RSS\_EXPANDED
- UPC\_A
- UPC\_E
- UPC\_EAN\_EXTENSION

# Stapelklasseneigenschaften

## OCR-Texterkennung

### Allgemeines

Die OCR Extraktion ist ein elementarer Teil der Squeeze Software. Dieser Kernbereich der Software ist mit verschiedenen Einstellungen versehen, die das Ergebnis der Extraktion tangieren. Im folgenden Artikel gehen wir auf die Besonderheiten und die Anforderungen der unterschiedlichen Eigenschaften ein.

#### Welche Arten der OCR unterstützt Squeeze?

Grundsätzlich unterscheiden wir im Kontext von Squeeze zwischen dem Einsatz einer OCR basierend auf den Ressourcen der lokalen Maschine und dem Einsatz eines Remote-OCR-Dienstes.

#### Was beinhaltet meine Standardversion von Squeeze?

Im Auslieferungszustand ist Squeeze mit einer lokal verfügbaren OCR-Engine ausgestattet. Auf Kundenwunsch können unsere Berater bei einer Squeeze Installation ab der Version 2.4 eine Remote-OCR aktivieren, die mithilfe von AI bessere Ergebnisse liefern kann.

## Allgemeine Stapelklassen-Eigenschaften

### OCREngine (ab Squeeze 2.4)

Wird diese Stapelklassen-Eigenschaft nicht konfiguriert greift automatisch die lokale OCR-Engine `ocrmypdf`.

Je nach Spezifikation und Lizenzierung ihres Squeeze-Systems können folgende Optionen für die OCREngine verwendet werden:

Squeeze Version	Optionen
-----------------	----------

ab 2.4.0	<code>default</code>
ab 2.4.0	<code>ai-ocr</code>
ab 2.5.0	<code>maxocr</code>
ab 2.6.0	<code>proxy-ocr</code>

## Voraussetzungen:

- default:
  - keine
- ai-ocr:
  - um die Remote-AI-OCR zu verwenden ist es notwendig dass eine Internetverbindung auf dem System existiert und dass die Anmeldedaten von Ihrem Squeeze Berater konfiguriert werden.
- maxocr
  - die konfigurierte Mandanten-Konfiguration/Server-Konfiguration für die [Dexpro Platform Integration](#).
  - die [MaxOCR konfiguration](#).
- proxy-ocr
  - die [KI Proxy Konfiguration](#)

# Stapelklassen-Eigenschaften für die lokale OCR Engine

## OCRForce

Im Standard wird bei digitalen PDF's der Textlayer genutzt und die Felderkennung darauf angewendet (`false`). Um aber eine OCR zu erzwingen ist dieser Schalter auf `true` zu setzen.

## OCRLanguage

Im Standard werden die Sprachpakete Deutsch und Englisch verwendet. Für die deutsche Detektion wird der Wert `deu` eingetragen und für die englische Detektion der Wert `eng` eingetragen.

Hier können projektspezifisch auch weitere Sprachen oder abgewandelte Sprachpaket-Varianten angegeben werden, bei denen die OCR schneller/langsamer bzw. mit niedriger/höherer Qualität Ergebnisse liefert. Im folgenden eine Übersicht über die im Standard enthaltenen Sprachpakete:

Squeeze Version	Optionen
vor 2.4.0	<ul style="list-style-type: none"> <li>• deu</li> <li>• eng</li> </ul>
ab 2.4.0	<ul style="list-style-type: none"> <li>• deu, deu_best, deu_fast, deu_std</li> <li>• lat_best, lat_fast, lat_std</li> <li>• eng</li> <li>• osd</li> </ul>

## OCRPageLimit

Anzahl der auszulesenden Seiten im Dokument. Syntax n-m

Beispiel für Auslesung der ersten 3 Seiten: 1-3

## PDFA-Conversion

Es wird ein PDFA kompatibles Dokument erzeugt. Eingabe 1|0 (☐true||☐false)

## PDFProcessor

Hier gilt ☐PDFBox als Standard. ☐PDFMiner ist die Alternative .

## PSM-Modes

Im Project bietet es sich an, die Modi 3, 4, 6 und 11 zu verwenden. Dabei gilt 3 als Standard.

3	Standardeinstellung liefert gute Ergebnisse.
4	Wortweise Segmentierung. Es wird nicht nach Zeilen geschaut sondern Worten. (verfügbar ab Version 2.0)
6	Gut für Positionsdaten. Hat aber Probleme bei Linien die sehr dicht am Text sind.
11	Gut bei vielen Grafiken auf den Dokumenten.

## OCRRotationThreshold

Mit dieser Eigenschaft können Sie beeinflussen wie *agressiv* Seiten in der OCR gedreht werden. Nutzen Sie diesen Wert, wenn Dokumente falsch gedreht werden.

Geringe Werte führen dazu, dass mehr Dokumente gedreht werden. Die Software muss sich also nicht sehr sicher sein, dass eine Seite rotiert werden muss.

Hohe Werte führen dazu, dass Dokumente seltener gedreht werden, also nur wenn sich die Software sehr sicher ist, dass eine Seite rotiert werden muss.

Im Standard ist dieser Wert `9.0`

# Stapelklassen-Eigenschaften für die Remote-AI-OCR/MaxOCR/KI-Proxy Engine

Aktuell gibt es keine Möglichkeiten die Remote-AI-OCR zu beeinflussen.

## Fragen und Antworten?

1. Ich habe die `ai-ocr/maxocr` als OCREngine Eigenschaft ausgewählt, jedoch funktioniert die Texterkennung nicht mehr ?
  - Gehen Sie bitte Sicher das Ihr Squeeze Berater die notwendigen Anmeldedaten zur Aktivierung der Remote OCR hinterlegt hat.
2. Ich habe mit der Remote-AI-OCR ein Dokument verarbeitet, mehrere Dokumente liefen erfolgreich durch, jedoch bleibt dieses Dokument hängen.
  - Aufgrund der begrenzten Ressourcen kann die AI-Remote-OCR maximal 100 Seiten pro Dokument verarbeiten. Überprüfen Sie daher die Anzahl der Seiten und nutzen bei nicht erfolgreicher Verarbeitung die lokale OCR.
3. Ich nutze die Remote-AI-OCR und mein Dokument hat mehrere Seiten jedoch nicht mehr als 100 Seiten trotzdem hängt das Dokument in der Verarbeitungskette fest.
  - Squeeze wartet insgesamt 3 Minuten auf die Verarbeitung des Dokumentes. Konnte der entfernte Dienst innerhalb dieser 3 Minuten das Dokument nicht verarbeiten, wird Squeeze eine Fehlermeldung mit einem Timeout Hinweis liefern. Schieben Sie das Dokument erneut über die technische Warteschlange in den Schritt "Texterkennung" Squeeze prüft in dem Fall ob das bereits hochgeladene Dokument verarbeitet wurde.