

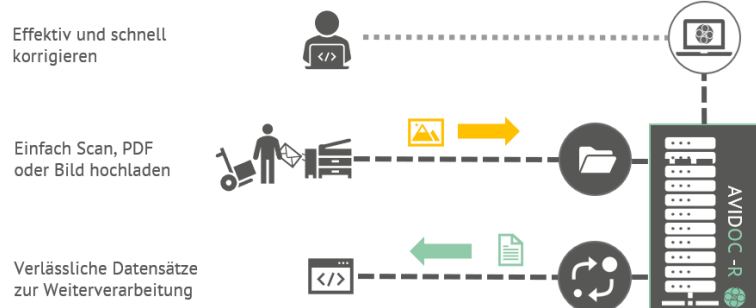


AVIDOC-R

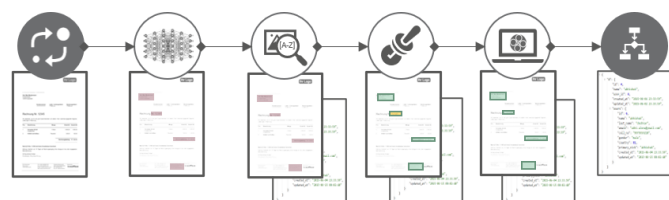
FUNKTIONSBESCHREIBUNG

AVIDOC-R: Einfach schnell digitalisieren

Egal ob Scan, Foto oder PDF – mit AVIDOC-R lassen sich relevante Informationen gezielt auslesen. Die automatische Validierung und Formatierung der ausgelesenen Daten sichert die Verlässlichkeit der Datensätze und unterstützt den Anwender darüber hinaus bei der schnellen Korrektur (Grab and Go). Hier gibt der AVIDOC-R stets das passende Feedback, wenn Daten einfach und schnell korrigiert und ergänzt werden. AVIDOC-R ist flexibel in der Annahme der Daten: AVIDOC-R kann Daten aus Laufwerken ziehen (pull), über eine REST-API annehmen oder ein Upload-Portal zur Verfügung stellen, um die Dokumente einfach hochzuladen.



AUFBAU einer AVIDOC-R Anwendung

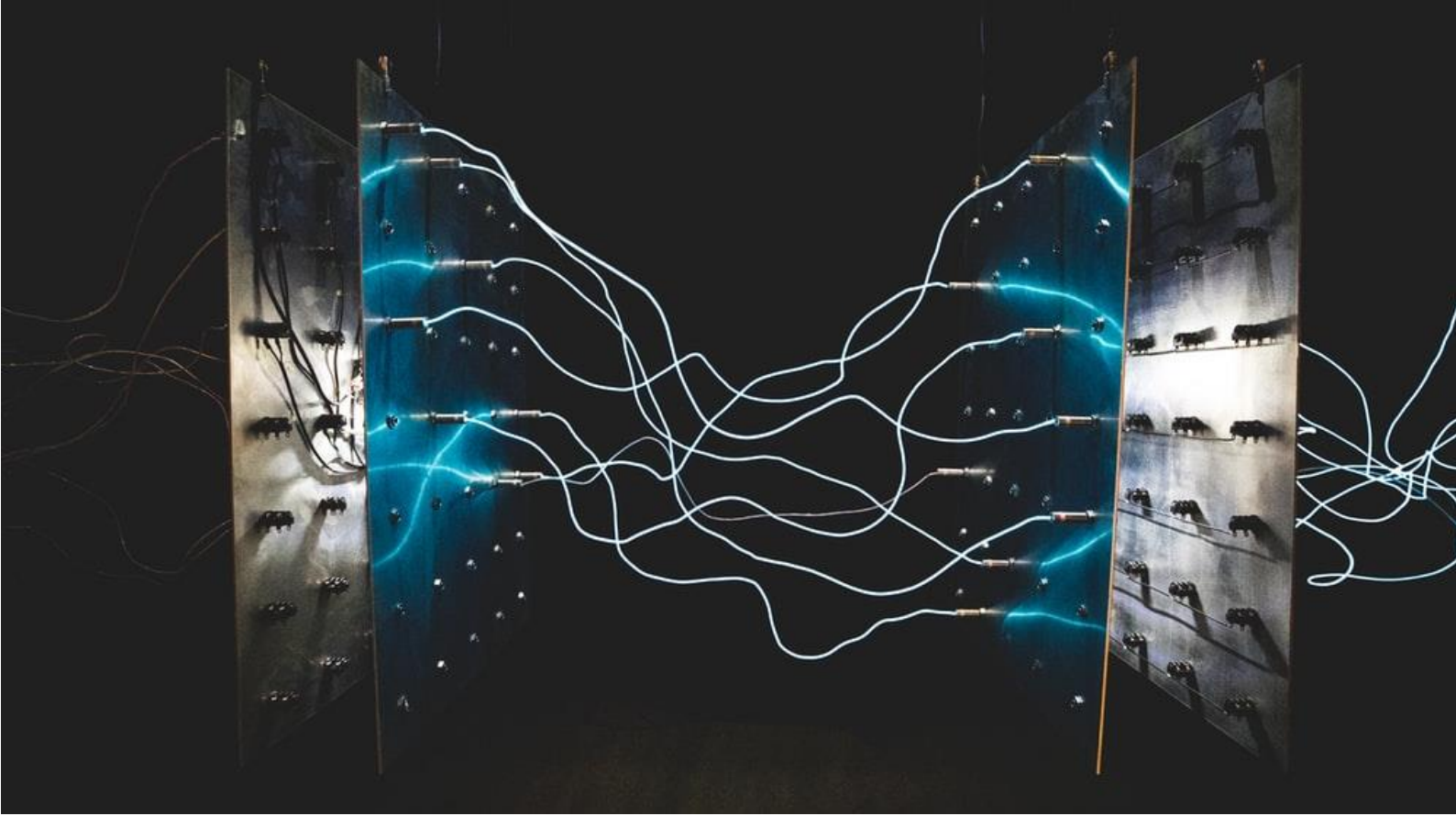


ABLAUF der Bilderkennung und Extraktion

AVIDOC-R arbeitet in sechs Schritten (s. Abbildung rechts): **Upload** der Daten, **Identifikation** der Datenfelder auf den Dokumenten, **Extraktion** der Texte, **Formatierung und Validierung**, Schnelle **Korrektur** und zeitnahe **Übersmittlung** der Ergebnisse. Drei Schnittstellen (Abbildung links) dienen dem Senden und Empfangen der Daten, eine Benutzerschnittstelle ermöglicht die einfache Korrektur, das Monitoring und die Administration des Systems.

Inhalt

AVIDOC-R: Einfach schnell digitalisieren.....	1
Schnittstellen, Datenupload und Rückgabe.....	2
Benutzerverwaltung und Authentifizierung.....	2
Qualitätskennzahlen und Statistiken	3
Formate und Mindestanforderungen.....	4
Grab and Go – Korrektur und Erfassung einfach beschleunigt	5
Funktionen der Prozessautomatisierung.....	6
Automatische Formatierung von Feldtypen und Validierung.....	7



Schnittstellen, Datenupload und Rückgabe

Datenannahme:

Die Bilddateien müssen in Ordnern (Stapeln) gruppiert zur Verfügung gestellt werden. Wird ein SFTP-Server verwendet, kann in AVIDOC-R eine **Signaldatei konfiguriert** werden, die dem System anzeigt, dass die Stapel vollständig sind und geladen werden können. AVIDOC-R ermöglicht ebenfalls den Upload der Stapel über SFTP in einem konfigurierten Ort auf den Servern der atacama blooms.

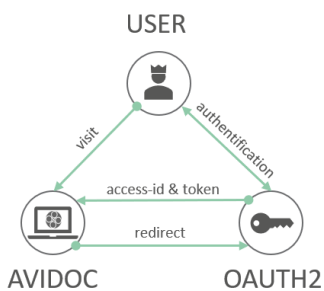
Datenrückgabe:

Für die Rückgabe der Ergebnisse können die Daten auf einen SFTP-Server geladen oder auf dem SFTP-Server der atacama blooms abgeholt werden. Die Dateien entsprechen der Eingangsstruktur und werden in einem Container abgelegt, der mit dem genauen Zeitpunkt der Erstellung bezeichnet ist.

Zeitgesteuerte Rückgabe: Im AVIDOC-R kann die Rückgabe einzelner Dokumente zeitgesteuert angefordert werden. In definierten Intervallen (z. B. alle 4 Stunden) werden bis dahin fertiggestellte Datensätze als komprimierter Container (ZIP-Dateiformat) im konfigurierten Ausgangsverzeichnis zur Verfügung gestellt.

Rückgabe vollständiger Stapel: Werden vollständige Stapel angefordert, werden die Ergebnisse eines vollständig verarbeiteten Stapels als komprimierter Container (ZIP-Dateiformat) im konfigurierten Ausgangsverzeichnis zur Verfügung gestellt.

Benutzerverwaltung und Authentifizierung



AVIDOC-R setzt für die Authentifizierung der Benutzer am „Grab and Go“, Monitoring oder den administrativen Oberflächen auf OAUTH2 und eine mächtige Benutzerverwaltung. Diese ermöglicht z. B. die Umsetzung von Rollenkonzepten und vieles andere mehr. Die Authentifizierung ist ein eigenständiger Applikationsserver, der separat von den abzusichernden Anwendungen läuft.

In der Abbildung ist zu sehen, wie AVIDOC-R jeden Benutzer zur Anmeldung an die Authentifizierung weiterleitet, welche die Credentials (Berechtigungsnachweise) entgegennimmt und weiterverarbeitet. Das Vorgehen isoliert die Anwendung von den zu schützenden Anmeldeinformationen der Benutzer. Das reduziert potenzielle Sicherheitslücken bei der Verarbeitung und Weiterleitung sensibler Credentials.

Eine **Zwei-Faktor-Authentifizierung** sichert den Zugang zur Applikation weiterhin durch einen zweiten Faktor ab. Mit einer speziellen App schützt die Anwendung so auch vor den Folgen von Phishing und sichert den Zugang gegen Brute-Force-Attacken und anderen Angriffsszenarien ab. So kann der Benutzer zuverlässig, sicher und gemäß den Richtlinien der DSGVO die Services nutzen.

Management und Monitoring

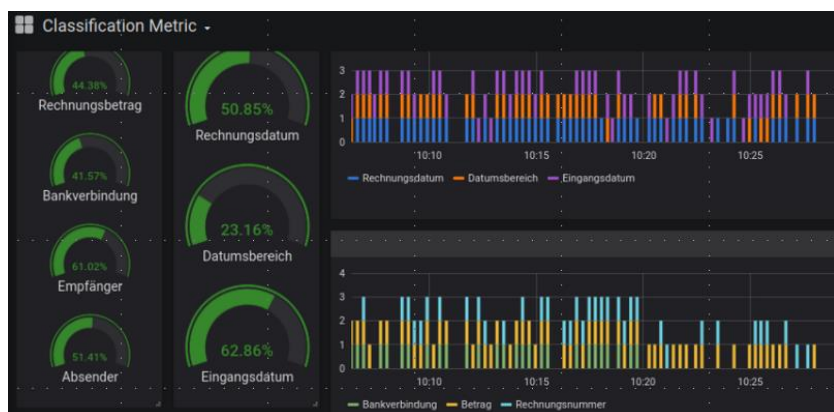
Der administrative Benutzer behält die Qualitätskennzahlen und Performance des Systems stets im Blick. Auf übersichtlichen Dashboards lassen sich die Anzahl der Datensätze, Erkennungs- und Validierungsquoten monitorieren.

Bei signifikanten Problemen können darüber hinaus jedes einzelne Dokument näher beleuchtet und die Teilergebnisse sowie Bearbeitungsschritte für jedes Feld drill-down betrachtet werden.

Qualitätskennzahlen und Statistiken

Das System AVIDOC-R stellt für Dunkelverarbeitung und schnelle Korrektur in einem übersichtlichen Dashboard einige Metriken zur Verfügung, die jederzeit einen Überblick über die Geschwindigkeit und Qualität der Verarbeitung geben:

Bearbeitungszeit je Dokument	Gibt an, wie viele Sekunden der AVIDOC-R für die Dunkelverarbeitung jedes einzelnen Dokuments benötigt. So kann rechtzeitig reagiert werden, wenn sich das Aufkommen der Dokumente erhöht, die Qualität der Eingangsdokumente sinkt oder die Zahl einzelner Seiten je Dokument steigt.
Bearbeitungszeit Warnung	Werden über einen längeren Zeitraum mehr Dokumente in die Warteschlange geladen als in der aktuellen Systemspezifikation verarbeitet werden können, wird eine Warnung angezeigt, auf die der administrative Anwender reagieren kann.
Bearbeitungszeit je Korrektur	Die durchschnittliche Zeit vom Laden/Anzeigen eines Dokuments bis zum Absenden der Korrektur. Diese Daten werden anonym über alle Nachbearbeitungsvorgänge erfasst.
Korrekt-Klassifikationsrate (KKR)	Quote der korrekten automatisierten Erfassung. Eine nähere Beschreibung lesen Sie im nächsten Abschnitt.



Die Korrekt-Klassifikations-Rate (KKR) wird für diejenigen Dokumente ermittelt, die alle Mindestanforderungen erfüllen. Die Berechnungen aller Qualitätskriterien beziehen sich zunächst auf die einzelnen Datenfelder. Die KKR einzelner Felder kann variieren. Bestimmte Felder müssen überdurchschnittlich häufig nacherfasst werden, während andere Felder nahezu vollständig automatisiert erkannt werden. Die schnelle Korrektur ist darauf spezialisiert, diese einzelnen Korrekturen in Sekunden zu erledigen, um so die Bearbeitungszeit je Dokument radikal zu verkürzen.

Korrekt klassifiziert ist ein Datenfeld, wenn die Mindestanforderungen erfüllt sind und das Datenfeld valide Daten enthält, konvertiert und validiert wurde (*true positive* oder *positive hit*) oder das Datenfeld KEINE validen Daten enthält und nicht konvertiert wurde (z.B. ungültige IBAN, IK-Nummer etc.) (*false positive* oder *correct rejection*).

Nicht korrekt ist ein Datenfeld klassifiziert, wenn die Mindestanforderungen erfüllt sind und das Datenfeld valide Daten enthält und nicht konvertiert und validiert wurde (*miss*) oder das Datenfeld KEINE validen Daten enthält, aber dennoch konvertiert und validiert wurde (z.B. ungültige IBAN, IK-Nummer etc.) (*fallout*).

Das Dokument ist insgesamt korrekt klassifiziert, wenn alle Felder, die valide Daten enthalten, auch als solche erkannt wurden. Diese Quote wird über das Produkt aller KKR der einzelnen Felder gebildet.

Formate und Mindestanforderungen

Unterstützte Bildformate:

Portable Network Graphics	Dateisuffix .png
Joint Photographic Experts Group	Dateisuffix .jpg, (visuell) verlustfreie Komprimierung nach ISO/IEC IS 10918-1, ISO/IEC IS 14495-1. Bei verlustbehafteter Komprimierung kann es zu geringeren Erkennungsquoten kommen.
Tagged Image File Format	Dateisuffix .tiff
Windows Bitmap	Dateisuffix .bmp
Device Independent Bitmap	Dateisuffix .dib

Die Bildformate können auch in ein PDF eingebettet werden. Hier ersetzt das PDF die Ordnerstruktur, die einzelne Bilder zu einem Dokument zusammenfasst (z. B. bei mehrseitigen Rechnungen)

Mindestanforderungen an die Qualität der Dokumente

- Notwendige Qualität des Eingangsbildes 300dpi, farbig
- Hinreichende Qualität min. 400dpi, farbig, verlustfreie oder keine Komprimierung
- Maschinengeschriebenes Dokument mit einer Schriftgröße von min. 8pt
- Alle als notwendig deklarierten Datenfelder sind auch vorhanden (z. B. Rechnungsnummer, Bankverbindung)
- Dokument ist vom eingestellten Typ/Klasse (z. B. Lieferschein, Rechnung)

Hinweis zu Qualitätskennzahlen bei ungenügenden Eingangsdokumenten

Ungenügend erfüllte Mindestanforderungen bedeuten nicht automatisch, dass das Dokument nicht verarbeitet werden kann. Eine Aussage über die Erkennungsqualität des AVIDOC-R lassen diese Dokumente jedoch nicht zu und so werden diese z. B. auch nicht für die Berechnung von Qualitätskennzahlen berücksichtigt. Hintergrund ist, dass sich die Qualitätskennzahlen zu Gunsten des Produkts verschieben könnten, denn die Aussortierung dieser Dokumente ist ein korrektes Verhalten des AVIDOC-R. Für Dokumente geringerer Qualität weist die atacama blooms gerne eine gesonderte Statistik und passgenaue Qualitätskennzahlen aus.

Technische Voraussetzungen

Client (Grab and Go, Administration, Monitoring)	Für den Zugriff auf AVIDOC-R werden eine Internetverbindung (min. 4Mbit) sowie eine aktuelle Version von Chrome oder Firefox benötigt. Der Zugriff mit anderen Browsern kann zu Fehlern in der Anzeige oder Bedienung führen. Für die Bearbeitung wird der Vollbildmodus des Browsers auf einem Full-HD Bildschirm empfohlen. Als Eingabegeräte werden eine Maus mit min. 4 Tasten empfohlen, sowie eine Tastatur mit Nummernfeld und Funktionstasten (F1-F12). Das Arbeiten an Laptops ist möglich, es wird jedoch empfohlen, eine externe Maus und Tastatur zu verwenden.
SFTP (Client/Server Alternative)	SFTP-Client: Das anfordernde System stellt eine Verbindung zum sicheren SFTP-Server des AVIDOC-R her. Die Daten werden dort abgelegt und verarbeitete Daten abgeholt. SFTP-Server: Das anfordernde System stellt einen Ordner für die sichere Übertragung der Daten über SFTP zur Verfügung.
Bandbreite Datenübertragung	Die Bandbreite des Uploads muss der Menge an Daten angemessen sein. Ein Stapel sollte in weniger als 10 Minuten übertragen werden können, um Fehler in der Übertragung zu vermeiden. Dies kann sowohl über eine Gewährleistung der Bandbreite für den Upload als auch durch eine Reduzierung der Stapelgröße erreicht werden.



Schnelle Korrektur

Für alle Fälle, in denen die automatische Erkennung nicht alle Datenfelder erkennt oder ein neuer Dokumenten Typ angelernt werden soll, hilft die einfach schnelle Korrektur „Grab and Go“.

Die intuitive und auf schnelle Eingabe optimierte Oberfläche lässt Sie kinderleicht kleinere Fehler in der Erkennung korrigieren und im gleichen Schritt den AVIDOC aus Ihren Eingaben lernen.

Grab and Go – Korrektur und Erfassung einfach beschleunigt

Vorbereitete Segmentierung, OCR-Unterstützung, Parsen und Formatieren von Adressen, Datum, Rechnungsnummern, Prüfung der IBAN und PLZ, Klassifikation „Art der Leistung“, Positives Feedback, intelligentes Nachladen vom Stapel. „Grab and Go“ kann als Website in Ihren Workflow integriert werden. Durch einen sicheren Aufruf betten Sie „Grab and Go“ einfach in Ihre Anwendung ein oder stellen eigene Terminals bereit, die auf die Nacherfassung spezialisiert sind.

Features

Auswahl der Daten durch Klicken und Ziehen	Einfache Auswahl bereits segmentierter Felder auf einem Dokument mit einem Klick.
Automatische Konvertierung	Die Datenfelder lassen stets nur Formate zu, die auch zum Datentyp passen. So kann der Benutzer die Eingaben schnell und ohne das Betätigen von Steuerzeichen (Komma, Punkt, Leertaste) erledigen.
Sichere Validierung	Die Eingaben werden sofort auf Plausibilität geprüft, so dass die Datenqualität schon vor der späteren logischen Auswertung so korrekt wie möglich übergeben werden kann.
Schnelle und ergonomisch optimiert erreichbare Funktionen	Alle Funktionen sind einfach über Schnell Tasten erreichbar. Diese sind nach ergonomischen Gesichtspunkten voreingestellt oder können beliebig konfiguriert werden.

Funktionen

Wechsel zwischen Datenfeldern	Zwischen den Datenfeldern des Obermenüs und der Untermenüs kann jederzeit gewechselt werden .
Wechsel zwischen unbearbeiteten Datenfeldern	Zwischen unbearbeiteten Feldern kann jederzeit gewechselt werden.
Hinzufügen von Datenfeldern	In einem Datenfeld können zusätzliche Einträge hinzugefügt werden.
Entfernen von Datenfeldern	Einträge in Datenfeldern können gelöscht werden.
Selektion von Informationen	Im Dokument können Informationen selektiert werden.
Validierung von Informationen	Selektierte Informationen werden direkt nach Auswahl überprüft und entsprechend farblich markiert: gelb: Information konnte nicht validiert werden. grün: Information konnte validiert werden.
Anzeige von Informationen	Selektierte Informationen können jederzeit angezeigt werden.
Anzeige von Informationen zur Bearbeitung	Der aktuelle Bearbeitungsmodus (Freihandselektion, Hilfsboxen) wird in einer Infobox angezeigt.
Bearbeitungszeit	Die Bearbeitungszeit ist begrenzt und wird in einer Infobox angezeigt.

Deep Learning

Mit AVIDOC-R setzt die atacama blooms stets auf moderne Ansätze der KI. Für die Analyse der Dokumente wird unter anderem „Deep-Learning“ verwendet. Dies ist eine Methode des maschinellen Lernens, die künstliche neuronale Netze mit zahlreichen Zwischenschichten zwischen Eingabeschicht und Ausgabeschicht einsetzt und dadurch eine umfangreiche innere Struktur herausbildet. So kann das System selbstständig herausfinden, wie das Layout eines Dokuments zu verstehen und der Inhalt einzuordnen ist.

Für die datenstromorientierte Implementierung der Informationsgewinnung werden Tensoren mit spezialisierten Hardwarekomponenten, sogenannten Tensor-Cores, auf spezialisierten Grafikkarten verknüpft. So werden sehr komplexen Rechenoperationen performant abgearbeitet. Vor allem das Training eines mehrschichtigen neuronalen Netzes erfordert hier ein besonderes Augenmerk.

Funktionen der Prozessautomatisierung

Stapelverarbeitung

Ein bereitgestellter Stapel von Dokumenten enthält mindestens ein Dokument. Die Größe der Stapel sollte an die Bandbreite des Uploads angepasst sein. Der Stapel bildet die kleinste Einheit an Daten, die in einem Schritt an den AVIDOC-R übermittelt wird.

Klassifikation des relevanten Dokumenten-Typs

AVIDOC-R erkennt anhand der Dokumentenstruktur und Datenfelder, welches der einzelnen Bilder eines Dokuments die geforderten Daten enthält. Enthält eine im Stapel als ein Dokument zusammengefasste Menge von Bildern mehrere Dokumente, versucht der AVIDOC-R diese zu trennen und erzeugt im JSON-Format eine Liste von Dokumenten.

Identifizierung der Datenfelder

AVIDOC-R ermittelt die Existenz und den Ort der Datenfelder und erstellt eine Menge von Kandidaten für jedes einzelne Feld. Die Menge der Kandidaten wird im Anschluss formatiert und validiert, so dass die gefundenen Felder mit hoher Wahrscheinlichkeit das geforderte Datum repräsentieren. Von vorneherein kann nicht ausgeschlossen werden, dass ein Datenfeld mehrfach auf einem Dokument auftaucht. Diese können identischen Inhalts sein (z. B. wird die KV-Nummer an verschiedenen Stellen des Dokuments erscheinen), es können aber auch mehrere Objekte auftauchen (z. B. mehrere Bankverbindungen eines Unternehmens).

Extraktion, Formatierung und Validierung

Der Inhalt der einzelnen Datenfelder wird ausgelesen und anhand der Vorgaben des Feldes transformiert. Hierbei versucht AVIDOC-R bereits plausible Korrekturen (z. B. des Ortes bei bekannter Postleitzahl, Centbeträge als „00“ bei glatten Rechnungsbeträgen) vorzunehmen. In mehreren Schritten werden die Datenfelder durch gegebene Regeln und Standards validiert. Eine genauere Beschreibung der verfügbaren Validierungen finden Sie im nächsten Abschnitt.

Konstruktion des Datenobjekts

Die erfassten Felder werden in ein JSON-Objekt geschrieben. Dann wird beurteilt, ob das Dokument vollständig ist oder nachbearbeitet werden muss. Im letzteren Fall wird dieses Objekt mit dem Dokument der schnellen Korrektur zur Verfügung gestellt.



Automatische Formatierung von Feldtypen und Validierung

AVIDOC-R stellt für verschiedene Datentypen automatische Formatierungen und Validierungen zur Verfügung. Formatierung heißt, dass unterschiedliche Schreibweisen von Informationen (z. B. 10,00 EUR, 10,-€, EUR 10,-) in eine maschinenlesbare Standardform umgewandelt werden. Systeme können so verlässlich mit einheitlichen Informationen weiterarbeiten.

Die Validierungen unterstützen dabei, offensichtliche Fehler in der Erkennung oder auf den Dokumenten selbst, zu entdecken. In der Schnellbearbeitung können diese einfach korrigiert werden. Auch hier wird der Benutzer mit diesen Funktionen bei der Eingabe unterstützt. So werden Eingabefehler vermieden und im Ergebnis weitgehend schlüssige Datensätze erzeugt.

Angeschlossene Systeme können den AVIDOC-R hier aktiv unterstützen. Jeder Datentyp kann mit Positiv- und Negativlisten verknüpft werden, die z. B. alle bekannten Kunden- oder Kreditorenummern enthalten, die ausschließlich automatisch verarbeitet werden, oder Rechnungspositionen, die auf keinen Fall weiterverarbeitet werden sollen.

Formatierungen und Validierungen:

Kalenderdatum	AVIDOC-R erfasst verschiedene Datumsformate (z.B. 01.05.20 oder 01. Mai 2020) und konvertiert diese in die Form dd.mm.jjjj, also z. B. den 01.05.2020. Validiert wird außerdem, ob das Datum existiert.
Kalenderzeitraum	Neben einzelnen Daten erfasst der AVIDOC-R Zeiträume unterschiedlicher Form, die z. B. Rechnungs- oder Leistungszeiträumen entsprechen können. Beispiele sind: von 01.05. bis 31.05.2020, Mai 2020, 01.05.2020-31.05.2020 und viele andere mehr.
IBAN	Die Internationale Bankkontonummer (englisch: International Bank Account Number, abgekürzt IBAN, ISO-Norm ISO 13616-1:2007 Teil 1). Wird eine IBAN aus Dokumenten ausgelesen oder durch den Anwender erfasst, wird diese für den digitalen Austausch formatiert und validiert.
Adresse	Postleitzahl und Ort werden auf ihre Gültigkeit und ihre Zuordnung validiert. Adressen werden in 3 Felder gespeichert: 1) Name inkl. Zusatz, 2) Straße oder Postfach und Nummer, 3) Postleitzahl und Ort.
Rechnungsbeträge	Beträge in Währungen werden in der Form *#,## ohne Tausendertrennzeichen und Währungsinformationen übergeben. Die Währung ist ein eigenes Datum.
IK-Validator Positivliste/Negativliste „IK-Klassen, IK-Nummern, IK-Regionen“	Die Institutionskennzeichen (kurz: IK) sind bundesweit eindeutige, neunstellige Zahlen, mit deren Hilfe Abrechnungen und Qualitätssicherungsmaßnahmen im Bereich der deutschen Sozialversicherung einrichtungsübergreifend abgewickelt werden können. Der AVIDOC-R verfügt sowohl über den Check der Prüfziffern als auch über einen Katalog der Einrichtungsklassen, gegen den die IK geprüft werden kann. Es sowohl können Positiv- als auch Negativlisten verwendet werden.
KV-Validator Parameter: „2012“ Positivliste „Versicherte“	Die Krankenversicherungsnummer dient der Identifikation des Versicherten bei einer Krankenversicherung: Ohne Parameter (freie Versicherten-Nummer): Prüft auf min. 6 Nummern oder KV-Nummer (s. u.). Parameter „2012“ : Prüft auf einen Großbuchstaben und 9 Ziffern sowie die Prüfsumme.
SGB XI §45b Positionen	AVIDOC-R erkennt die Positionen, die einer Leistung aus dem §45b entsprechen. Diese können mit den unterschiedlichen Leistungsarten verknüpft werden.
Rechnungsnummer	Rechnungsnummern folgen zwar gewissen rechtlichen Anforderungen, können aber sehr unterschiedlich aussehen. Daher kommt es bei der automatischen Erkennung zunächst oft zu Fragmenten, wie z. B. „Rechnung:“ oder „Nr.“, die AVIDOC-R zuverlässig entfernt und die Rechnungsnummer auf ihren tatsächlichen Wert ohne Leerzeichen konvertiert.



atacama blooms GmbH & Co. KG
Universitätsallee 15
28359 Bremen

Fon. +49 (0) 421 – 365 122 40
Fax. +49 (0) 421 – 223 82 85
Mail. info@atacama-blooms.de
Web. www.atacama-blooms.de

