

## INHALTSVERZEICHNIS

Automatische Erkennung von Rechnungen und Belegen .....	2
Das Angebot.....	2
Datenschutz .....	2
Probleme beim Scannen von Rechnungen .....	2
PDF-Rechnungen verarbeiten .....	3
Problem Splitt-Rechnungen .....	6
Automatisierte Erfassung als Alternative .....	6
Preisbetrachtung .....	7
Hintergrund .....	7
Weiterführende Informationen .....	8

## AUTOMATISCHE ERKENNUNG VON RECHNUNGEN UND BELEGEN

### DAS ANGEBOT

In den letzten Jahren sind eine Reihe von Anbietern am Markt erschienen, die die vollständige Automatisierung der Rechnungs- und Belegerfassung als Vorstufe für Rechnungseingangsbücher, Reisekostenerfassung und die Buchführung sowie das Banking anbieten.

Die Palette der Anbieter und Angebote ist vielfältig. Da gibt es Angebote der Banken – etwa die Deutsche Bank -, die in ihren Banking-Apps das Fotografieren und automatische Erfassen von Rechnungen anbieten, um den Rechnungsbetrag schnellstmöglich überweisen zu können. Da tritt natürlich Google auf den Plan und bietet mit Google Cloud Vision neben der Gesichtserkennung und der Bildersuche ebenfalls eine OCR (Optical Character Recognition – optische Zeichenerkennung) Funktion für Scans an. Dann gibt es Spezialanbieter wie Veryfi, Klippa oder taggun.io, die sich auf die reine Erkennung von Textinhalten in Dokumenten spezialisieren und hier vor allem Rechnungen und Belege im Blick haben – entsprechende Schnittstellen und Automatisierungen mit Buchhaltungsprogrammen werden angeboten. Und schließlich halten es einige Anbieter von Buchhaltungs-Software inzwischen für unabdingbar, den Kunden eine Vollautomatisierung der Buchhaltung anzubieten. Dank automatischer Verarbeitung von Eingangsrechnungen soll sich so die Buchhaltungsarbeit von selbst erledigen – so der Werbe-Tenor.

Warum macht INTex Publishing das nicht auch oder bietet zumindest Schnittstellen an?

### DATENSCHUTZ

Zunächst stellt sich die Datenschutz-Frage – auch Rechnungen enthalten personen-bezogene Daten, zumindest dann, wenn es um Ausgangsrechnungen an private Endverbraucher geht. Aber auch bei Einkäufen Ihrer Mitarbeiter für den persönlichen Arbeitsbedarf (etwa bei Bahntickets, Hotelrechnungen etc.) sind personen-bezogene Daten enthalten. Bei einem Scan und automatisierter OCR-Bearbeitung der Rechnung geben Sie indirekt diese Daten an einen Dritten weiter, setzen also auf Auftragsdatenverarbeitung. Und hierfür müssen Sie die Vereinbarkeit des Angebots mit der DSGVO sicherstellen.

### PROBLEME BEIM SCANNEN VON RECHNUNGEN

Nächster Schritt ist die Digitalisierung der Rechnungen. Bevor überhaupt an eine OCR-Verarbeitung gedacht werden kann, müssen die Belege erst mal digital vorliegen. Natürlich gibt es immer mehr Rechnungen, die man im PDF Format per Email erhält, aber weiterhin werden viele Lieferanten vom Handwerker über die Tankstelle bis hin zum Restaurant gedruckte Belege aushändigen. Diese müssen für die elektronische Weiterverarbeitung erst gescannt werden, was natürlich mit Aufwand verbunden ist. Ein Aufwand übrigens, den man sich bei der einfachen manuellen Erfassung etwa der Tankkosten ersparen kann.

Neben dem Aufwand für den Scan-Vorgang sind dann auch noch einige Hürden des Alltags zu überwinden. Die Scan-Software kann in der Regel für den OCR Vorgang nämlich nur dann ein vernünftiges Lese-Ergebnis bereitstellen, wenn die zu scannende Vorlage stimmt. Dagegen steht leider oft:

- Schief eingezogene Blätter am Scanner
- schräge Smartphone Fotos mit schlechter Belichtung und Unschärfe (bei der Arbeit mit Programmen wie OfficeLens)
- Falze, Knicke und Eselohren in der Papiervorlage
- Unsauberer Druck der Vorlage
- Kaffeeflecken
- Vergilbtes und aufgerolltes Thermopapier

## Probleme der automatischen Rechnungserkennung – INTex Publishing

Auf Basis dieser Alltagsprobleme fiel ein Test, den wir mit der Deutschen Bank Smartphone Banking-App gemacht haben, sehr ernüchternd aus. Von vier getesteten Rechnungen konnte nur eine Rechnung korrekt gelesen werden, d.h. Bankverbindung, Rechnungsbetreff und Betrag waren richtig. Die korrekt erkannte Rechnung war übrigens eine Handy-Rechnung von 1&1- hier kann man vermuten, dass besondere Algorithmen für diese sehr gängige Rechnungsform zum Einsatz kamen. Bei der Rechnung eines örtlichen Handwerkers war das Testergebnis desaströs.

*Hinweis: OCR Marktführer Abbyy versteckt in den Support-Dokumenten den nicht unwichtigen Hinweis, dass für gute Ergebnisse im Zweifel ein „Anlernen“ der Software an den jeweiligen Rechnungstyp notwendig sein könnte. Das lohnt sich aber natürlich wiederum nur, wenn ein Rechnungsaussteller häufiger Rechnungen verschickt.*  
[https://help.abbyy.com/en-us/flexcapture/12/invoice\\_reader/ir\\_admintraining](https://help.abbyy.com/en-us/flexcapture/12/invoice_reader/ir_admintraining)

## PDF-RECHNUNGEN VERARBEITEN

Es ist wegen der Probleme mit den Scans vermutlich ohnehin ratsamer, die automatisierte Bearbeitung von Rechnungen auf die Rechnungen zu beschränken, die a) von vornherein digital erstellt wurden und b) als PDF-Datei vorliegen.

Wir haben hier einen Test mit der Software des Anbieters Veryfi gemacht. Zum Test dienten eine Rechnung des DSL Anbieters NetCologne, eine Rechnung des Website-Providers All-Inkl.com und eine unserer Rechnungen, die wir mit der Shop-Software Shopware erstellen.

Und das sind die Ergebnisse:

## NETCOLOGNE

The screenshot shows the Veryfi website interface. At the top, there is a navigation bar with the Veryfi logo, a search icon, and links for SOLUTIONS, PRICING, BLOG, RESOURCES, SIGN IN, and a FREE TRIAL button. Below the navigation bar is a 'Try It Now' section with a dashed blue border containing a 'Drag a receipt here to see a live demo' instruction and supported file types (JPEG/JPG/PNG/GIF images and PDF documents). The main content area displays a scanned invoice from NetCologne for February 2020. The invoice is partially processed, with some fields highlighted in yellow. To the right of the invoice, a metadata panel shows extracted information such as Vendor name, Address, Phone, Vendor type, Category, Payment, Invoice #, Date, and Due Date. Below this, a table lists the invoice lines with their descriptions and amounts in EUR. The total amount is EUR 29.90. At the bottom right, there is a feedback prompt: 'Are you satisfied with the results?' with thumbs up and down icons.

**NetCologne**  
Rechnungsdatum 04.02.2020  
Rechnungszeitraum 01.02. - 29.02.2020  
Bitte immer anrufen  
Rechnungsnummer 0012678116-2020  
Seite 1 von 2  
Bei Rückfragen Tel: 0221 2222-800 Fax: 0221 2222-300  
info@netcologne.de  
Vertragsnr. 0012678116-2020

Ihre Rechnung für Februar 2020

Leistungen	Anzahl	Grundpreis in EUR	Dauer (KMS) / Volumen (MB)	Gesamtbetrag in EUR	USt
<b>Doppel-Flat</b>					
Vertragsgültig: 30.06.2016					
Ende der aktuellen Vertragslaufzeit: 31.10.2020					
Kündigungsfrist: 3 Monate zum Laufzeitende					
Sonderer Kündigungsgangstermin: 31.07.2020					
<b>Monatliche Beträge</b>	1	29.9000		29.9000	19%
Doppel-Flat	1	-10.0000		-10.0000	19%
Aktionsvorteil 24 Monate Rabatt auf die Grundgebühr	1	10.0000		10.0000	19%
Bandbreite 100.000	1	0.0000		0.0000	19%
Mail and Cloud (ComCenter Basic)	1	0.0000		0.0000	19%
<b>Summe Doppel-Flat</b>				<b>29.9000</b>	

**Gesamtaufstellung der Rechnungsbeträge**

Zwischensumme (Netto)	25.1261	25.1261
Umsatzsteuer 19 % auf		4.7739
<b>Zu zahlender Betrag in EUR (Brutto)</b>		<b>29.90</b>

Der Rechnungsbetrag ist sofort fällig. Wir werden den Betrag in Höhe von EUR 29,90 abzüglich eines eventuellen Guthabens aus den Vormonaten am 15.02.2020 von dem hier von Internetprovider NetCologne angegebene Bankverbindung einziehen.  
Bitte SEPA. Bei einem vom Rechnungsempfänger abweichenden Kontoinhaber stellen Sie bitte sicher, dass der Kontoinhaber rechtzeitig von dieser Vorkündigung in Kenntnis gesetzt wird.

NetCologne Zentrale Köln

**Vendor name:** NetCologne Zentrale Köln  
**Address:** Am Coloneum 9, 50829 Köln, Germany  
**Phone:** 0221 2222-800  
**Vendor type:** point\_of\_interest, store, establishment  
**Category:** Utilities  
**Payment:** \*\*\*3993  
**Invoice #:** 0012678116-2020  
**Date:** 2020-03-04 00:00:00  
**Due Date:** 2020-03-04

Line	Description	Amount (EUR)
Line 1:	Monatliche Beträge	EUR 0.00
Line 2:	Doppel-Flat	EUR 29.90
Line 3:	Aktionsvorteil 24 Monate Rabatt auf die Grundgebühr	EUR -10.00
Line 4:	Bandbreite 100.000	EUR 10.00
Line 5:	Mail and Cloud (ComCenter Basic)	EUR 0.00
Line 6:	Gesamtaufstellung der Rechnungsbeträge	EUR 0.00
Tax 1:	Tax (19%)	EUR 4.77
Subtotal:		EUR 25.13
Tax:		EUR 4.77
Tip:		EUR 0.00
Total amount:		EUR 29.90

Are you satisfied with the results?

## Probleme der automatischen Rechnungserkennung – INTex Publishing

Ergebnisse sind recht gut, Adresse und Rechnungsnummer sowie Rechnungsdatum und Gesamtbetrag wurden richtig erkannt.

Falsch ist die Interpretation einer Zwischenüberschrift „Monatliche Beträge“ als Positionszeile a 0.- Euro. Nicht ganz korrekt ist auch die von Netcologne eigentlich mit vier Stellen nach dem Komma ausgewiesene Netto- und Steuersumme als gerundete Beträge – dies ist etwas eigenmächtig, aber hinnehmbar.

Insgesamt ein gutes Ergebnis, das aber mit Upload und Verarbeitung ähnlich viel Zeit verschlingt, wie die manuelle Erfassung der Rechnung, sofern der Rechnungsaussteller bereits angelegt ist.

## ALL-INKL.COM

The screenshot shows the ALL-INKL.COM website. At the top, there is a navigation bar with 'SOLUTIONS', 'PRICING', 'BLOG', 'RESOURCES', a search icon, 'SIGN IN', and a 'FREE TRIAL' button. Below the navigation is a 'Try It Now' section with a dashed border and a blue cloud icon. The text says 'Drag a receipt here to see a live demo' and 'Supported files: JPEG/JPG/PNG/GIF images and PDF documents'.

The screenshot shows the ALL-INKL.COM invoice and its JSON representation. The invoice is from 'Neue Medien Münnich' to 'INTex Publishing GmbH & Co. KG'. The invoice number is '02742' and the date is '27.02.2020'. The total amount is '269.46 €'. The JSON representation shows the following data:

```
{
  "Vendor name": "ALL-INKL.COM - Neue Medien Münnich",
  "Address": "Hauptstraße 68, 02742 Neusalza-Spremberg, Germany",
  "Phone": "35872 353",
  "Vendor type": "point_of_interest, establishment",
  "Category": "Other ***8504",
  "Payment": "D43",
  "Invoice #": "02742",
  "Date": "2020-02-27 00:00:00",
  "Due Date": "2020-02-27 00:00:00",
  "Line 1": "Grundgebühr EUR 299.40",
  "Line 2": "Vorauszahlungsrabatt EUR -299.40",
  "Tax 1": "MwSt (19%) EUR 43.02",
  "Subtotal": "EUR 226.44",
  "Tax": "EUR 43.02",
  "Tip": "EUR 0.00",
  "Total amount": "EUR 269.46"
}
```

Auch die Rechnung von All-Inkl. wird nicht wirklich schlecht erkannt.

Eigenmächtig ist die Software bei der Adresse des Rechnungsausstellers. Hier sorgt eine hinterlegte PLZ-Datenbank wohl dafür, dass aus der Rechnungsangabe 02742 Friedersdorf ein 02742 Neusalza-Spremberg wird. Das ist postalisch nicht wirklich falsch, steht aber so nicht auf der Rechnung.

Weiterhin wird die Kundennummer mit der Rechnungsnummer verwechselt.

Beim Vorauszahlungsrabatt passiert ein echter Patzer. Hier wird die Kommastelle verschoben und statt 10% Rabatt gibt es so satte 100% Rabatt.

The screenshot shows the Veryfi website interface. At the top, there is a navigation bar with 'SOLUTIONS', 'PRICING', 'BLOG', and 'RESOURCES' menus, along with a search icon and 'SIGN IN' and 'FREE TRIAL' buttons. Below the navigation bar is a 'Try It Now' section with a dashed blue border containing a cloud upload icon and the text 'Drag a receipt here to see a live demo'. Below this, it says 'Supported files: JPEG/JPG/PNG/GIF images and PDF documents'. The main content area is split into two parts. On the left is a preview of a scanned invoice from 'Intex Publishing GmbH & Co. KG'. The invoice includes a logo, contact information, and a table of items. The table has columns for 'Pos.', 'Art.Nr.', 'Bezeichnung', 'Art.', 'MwSt.', 'Brutto Preis', and 'Brutto Gesamt'. The total amount is 129.00 €. On the right is a 'Pretty | Raw JSON' view of the invoice data. It lists fields like 'Vendor name', 'Address', 'Phone', 'Vendor type', 'Category', 'Payment', 'Invoice #', 'Date', and 'Due Date'. It also shows a table of 'Line 1' and 'Tax 1' with their respective amounts in EUR.

Die von uns erstellte Standard-Rechnung wird – aus für uns nicht nachvollziehbaren Gründen – am schlechtesten erkannt. Obwohl unsere Adresse sogar zweifach ausgeworfen wird, ist diese überhaupt nicht erkannt worden. Automatisiert wäre also nicht klar, woher die Rechnung kommt.

Dafür stimmen diesmal die Beträge.

Fazit: Die Ergebnisse sind auf Basis digitaler PDF nicht schlecht, sogar deutlich besser als mit gescannten Druckvorlagen. Das war aber auch nicht anders zu erwarten. Fakt ist aber auch, dass keine einzige Rechnung völlig fehlerfrei erkannt worden ist. Großzügiger betrachtet war eine Rechnung korrekt, eine Rechnung fehlerhaft und eine unvollständig. Wenn das aber das Ergebnis ist, dann führt um eine manuelle Sichtung und Korrektur der automatisiert eingelesenen Rechnungen kein Weg herum. Die angedachte Effektivitätssteigerung verpufft auf diesem Wege aber schnell.

Ohnehin muss man auch klar sehen, dass die Software zwar im besten Fall eine Rechnung einlesen kann, aber eine inhaltliche Prüfung etwa auf die Frage hin, ob bestimmte Rechnungspositionen erbracht und fällig sind, kann die Software nicht leisten.

Genauso wenig kann die Automatisierung der Rechnungserfassung dafür sorgen, dass Rechnungen etwa Kostenstellen oder Projekten zugeordnet werden. Die digitale Entzifferung von Skonto-Bedingungen und Zahlungsbedingungen ist kaum möglich. Die Kennzeichnung des Zahlstatus als Offen oder Bezahlt kann ebenfalls nur durch Menschenhand erfolgen.

## PROBLEM SPLITT-RECHNUNGEN

Ein viel größeres Problem stellen allerdings die Rechnungen dar, die aus inhaltlichen oder buchhalterischen Gründen nach unterschiedlichen Kriterien zu splitten sind. Solche Rechnungs-Splitts können nämlich von den Programmen und vom OCR-Prozess in aller Regel überhaupt nicht erkannt und erfasst werden, da sie auf inhaltlichen und logischen Kriterien beruhen, die nicht per OCR auslesbar sind oder den Programmen unbekannt sind.

Beispiele:

- Spaltung von Einkaufsrechnungen nach Kostenstellen, etwa Toner-Kartuschen für die Abteilung/Kostenstelle Buchhaltung und Marker für das Marketing
- Spaltung nach MwSt-Sätzen, evtl. auch gemischt 7 und 19%, weil die Lieferung Bücher (7%) und Büromaterial (19%) enthielt.
- Spaltung nach Warengruppen, Angebots-Bestandteilen etc.

Gerade auch im Bereich der Hausverwaltung kommen weitere Splitts hinzu, die die Rechnungserkennung nur wenig hilfreich erscheinen lassen, weil sie damit nicht zurecht kommt:

- Spaltung von Handwerker-Rechnungen nach Arbeit und Material für den §35a EStG Ausweis von haushaltsnahen Dienstleistungen.
- Aufteilung eines Grundsteuer-Bescheides für ein Flurstück auf mehrere auf dem Grundstück befindliche Häuser
- Aufteilung von Hausmeister-Rechnungen in Gartenpflege, Treppenhausreinigung und Winterdienst
- Aufteilung von Handwerker-Rechnungen in umlegbare (Wartungen) und nicht umlegbare Bestandteile (Reparaturen)
- Aufteilung z.B. von Versicherungsrechnungen nach dem Leistungsprinzip auf zwei Abrechnungsjahre.
- Aufteilung z.B. einer Wasserrechnung zur Umlage nach Quadratmeter (Grundkosten) und Wasseruhren (Verbrauchskosten)

## AUTOMATISIERTE ERFASSUNG ALS ALTERNATIVE

Und deshalb setzen wir bei INTex Publishing mehr auf die Automatisierung der manuellen Erfassung als den Versuch, den menschlichen Verstand und die Augenprüfung einer Rechnung zu ersetzen. Und es gibt eine Menge Dinge, die man zur vereinfachten Erfassung tun kann:

- Einmalige Erfassung der Rechnungsaussteller (Kreditoren). Diese können auch etwa im Excel-Format importiert werden. Somit muss bei der Erfassung von Rechnungen meist nur der Kreditor aus einer Liste ausgewählt werden – die Adresse ist dann immer korrekt und zentral korrigierbar.
- Befüllung vieler Felder mit Vorgabewerten, z.B. Datum des Rechnungseingangs
- Autofill, Auswahl-Kalender und Vorgabelisten z.B. für Zahlungskonditionen
- Automatische Wiederholung von wiederkehrenden Rechnungen etwa für Abos und Abschläge durch die Datenbank selbst
- Muster-Funktion – so kann jede Rechnung als Vorgabe für eine weitere Rechnung dienen, etwa dann, wenn ein Kreditor zwar inhaltlich unterschiedliche Rechnungen ausstellt, aber eben der Kreditor, Zahlungsziele, Skonti etc. immer gleich sind.

Nebenher kann bei der manuellen Erfassung auf die möglicherweise schwierige Erstellung eines lesbaren, sauberen Scans von Papierrechnungen verzichtet werden.

Arbeitsvorgang	Mit Automatisierung durch OCR	Ohne OCR (INTex)
----------------	-------------------------------	------------------

<b>Scan von Papiervorlagen</b>	Unabdingbar in möglichst hoher Qualität	<b>Nicht notwendig</b>
<b>Upload oder Weiterleitung der digitalen Rechnung</b>	Unabdingbar für die Bearbeitung (DSGVO bedenken)	<b>Nicht notwendig</b>
<b>Erfassung des Kreditors</b>	Wird aus der Rechnung ausgelesen, u.U. immer wieder -> doppelte Datenhaltung	<b>Einmalig zentrale Erfassung</b>
<b>Erfassung der Rechnungsdaten wie Kundennummer, Rechnungsnummer</b>	Wird aus der Rechnung ausgelesen, manuelle Korrektur u.U. notwendig	Manuelle Erfassung mit Autofill Musterbuchungen
<b>Erfassung der Skonti und Zahlungsbedingungen</b>	Die OCR Erfassung bedeutet keine logisch-inhaltliche Erkennung	Berechnung von Skonto und Zahlungszieltag
<b>Inhaltliche Bearbeitung und Kategorisierung</b>	Nicht möglich -> manuelle Nacharbeit	Direkt bei der Erfassung manuell, automatisiert über Musterbuchungen
<b>Erfassung der Rechnungssummen</b>	Wird aus der Rechnung ausgelesen, manuelle Korrektur u.U. notwendig	Manuelle Erfassung
<b>Splittung etwa nach Kostenstellen</b>	Manuelle Erfassung, oft aber nicht möglich oder nicht vorgesehen	Manuelle Erfassung
<b>Erfassung der Positionszeilen</b>	Wird aus der Rechnung ausgelesen, manuelle Korrektur u.U. notwendig	Für die Weiterverarbeitung etwa zum Bezahlen der Rechnung oder Verbuchen in der Buchführung gar nicht notwendig
<b>Rechnungsprüfung (formal und inhaltlich)</b>	Manuelle Nachbearbeitung	Manuell
<b>Bearbeitungs- und Zahlstatus</b>	Manuelle Nachbearbeitung	Manuell
<b>Abo-Rechnungen</b>	Genauso viel Aufwand wie normale Rechnungen	<b>Vollautomatisierung</b>
<b>Ähnliche Rechnungen</b>	Gleicher Aufwand	<b>Muster-Rechnungen</b>

#### PREISBETRACHTUNG

Nicht außer Acht lassen sollten Sie auch die Kostenfrage. Die Rechnungserfassung durch Software-Dienstleister ist nämlich ebenfalls nicht kostenlos und die Anbieter mit integrierten OCR-Features lassen sich diese gut bezahlen. Während INTex Rechnungseingang PLUS im Jahr 79.- Euro incl. MwSt. kostet, verlangt z.B. Klippa ab 40.- Euro pro Monat.

Das Einlesen einer Rechnung mit Google Vision kostet 1,50 Dollar pro Rechnung, die ersten 1000 Rechnungen sind kostenlos. Taggun dagegen verlangt 8 US-Cent pro Rechnung, kommt mit seinem Sitz in Neuseeland aber vermutlich eher nicht als DSGVO-konformer Anbieter in Betracht.

#### HINTERGRUND

Abschließend wollen wir Ihnen noch erläutern, warum sich Software eigentlich so schwer mit dem Lesen von Rechnungen selbst aus digital erstellten PDFs tut und das Ganze nicht einfacher geht. Bei schlechten Scans und deren digitaler Umsetzung wird man das Problem ja noch nachvollziehen können, aber bei digitalen Vorlagen?

Hintergrund ist, dass das PDF-Dateiformat in seinem Ursprung kein Datenaustauschformat im technischen Sinne ist, also nichts mit CSV-Dateien, Excel-Tabellen oder gar XML-Dateien gemein hat. Diese Dateiartern



[XRechnung ...](#)

[ZugPferd ...](#)