



# Intelligent Indexing goes AI

Marc Hegemann

27.10.2022

# DocuWare



400  
Mitarbeiter



33  
Jahre im Geschäft



5  
Niederlassungen



15.000  
Kunden



100+  
Länder



## Wechsel in die Cloud

Mit Cloud-Software sind Skalierung, Leistungssteigerung und Mobilität für alle Teams in Reichweite. Und die Vorteile gegenüber traditioneller On-Premises-Software sind vielfältig: Finanzielle Flexibilität, unkomplizierte Verwaltung und IT, größere Akzeptanz. Mit [DocuWare Cloud](#) stehen Ihnen alle Funktionen der kompletten Plattform zur Verfügung. Ohne Kompromisse.



## Weniger Papier, mehr Umweltschutz

Jahr für Jahr verschlingt das Verwalten von Papierdokumenten Millionen: Unterlagen werden ausgedruckt, aufbewahrt, gesucht und manchmal wiederhergestellt werden – ein unerwünschter Ressourcenverbrauch. Digitalisierung [sichert nicht nur Ihre Daten](#) so, dass Sie jederzeit darauf zugreifen und Compliance-Anforderungen erfüllen können, sondern verbessert auch Ihre Ökobilanz.



## Automatisierung von Prozessen

Wenn ein digitaler Workflow automatisiert wird, profitiert davon jedes Teammitglied. Die Reduzierung manueller Schritte beschleunigt Genehmigungen. Der Wegfall von Dateneingaben minimiert die Fehlerzahl. Und durch Aufgabenerinnerungen bleiben alle am Ball. DocuWare liefert die [Tools, um komplexeste Workflows ohne zeitaufwendige Implementierung zu automatisieren](#).

Erste Schritte...

# Ressourcen-Aufbau



Personal – Suche nach einem AI Engineers



Bessere Identifizierung von Zielkandidaten/innen dank appliedAI



Wissensaufbau durch NLP und MLOps Trainings

# Use Case

***„Ermögliche dem User das automatische Indexieren von Dokumenten mit viel Fließtext“***

Beispiele:

- Verträge
- Angebote
- Emails
- Briefe & Anschreiben
- Wissenschaftliche Dokumente



# Intelligentes Indexieren von Rechnungen... 😊

The screenshot displays the DocuWare interface with a document form on the left and a scanned invoice on the right. Red arrows indicate the mapping of metadata fields to the invoice content.

**Document Form Fields:**

- Store to "Test Dokumentenpool"
- Art des Dokumentes \* Rechnung (eingehend)
- Firma \* HUGO:NET GmbH
- Datum \* 10.06.2008
- E-Mail \* info@hugonet.com
- Kontaktperson
- Kundennummer 85-41225
- Betreff \* RECHNUNG
- Status Genehmigt
- Fälligkeitsdatum 24.06.2008
- Dokumentnummer 349
- Projekt
- Betrag 37.033,17
- Steuernummer ID.No.: DE 987
- IBAN DE27 7008 0000 0123 4567 00
- Customer Service Rep genehmigt
- Tabellenfeld: 0 rows

**Invoice Content (HUGO:NET):**

\*\*\* HUGO:NET GmbH, Weinweg 7, D-71032 Bietigheim \*\*\*  
PETERS ENGINEERING  
THERESE-GIESE PLATZ 2  
82110 GERNERING  
GERMANY

\*\*\* RECHNUNG \*\*\*

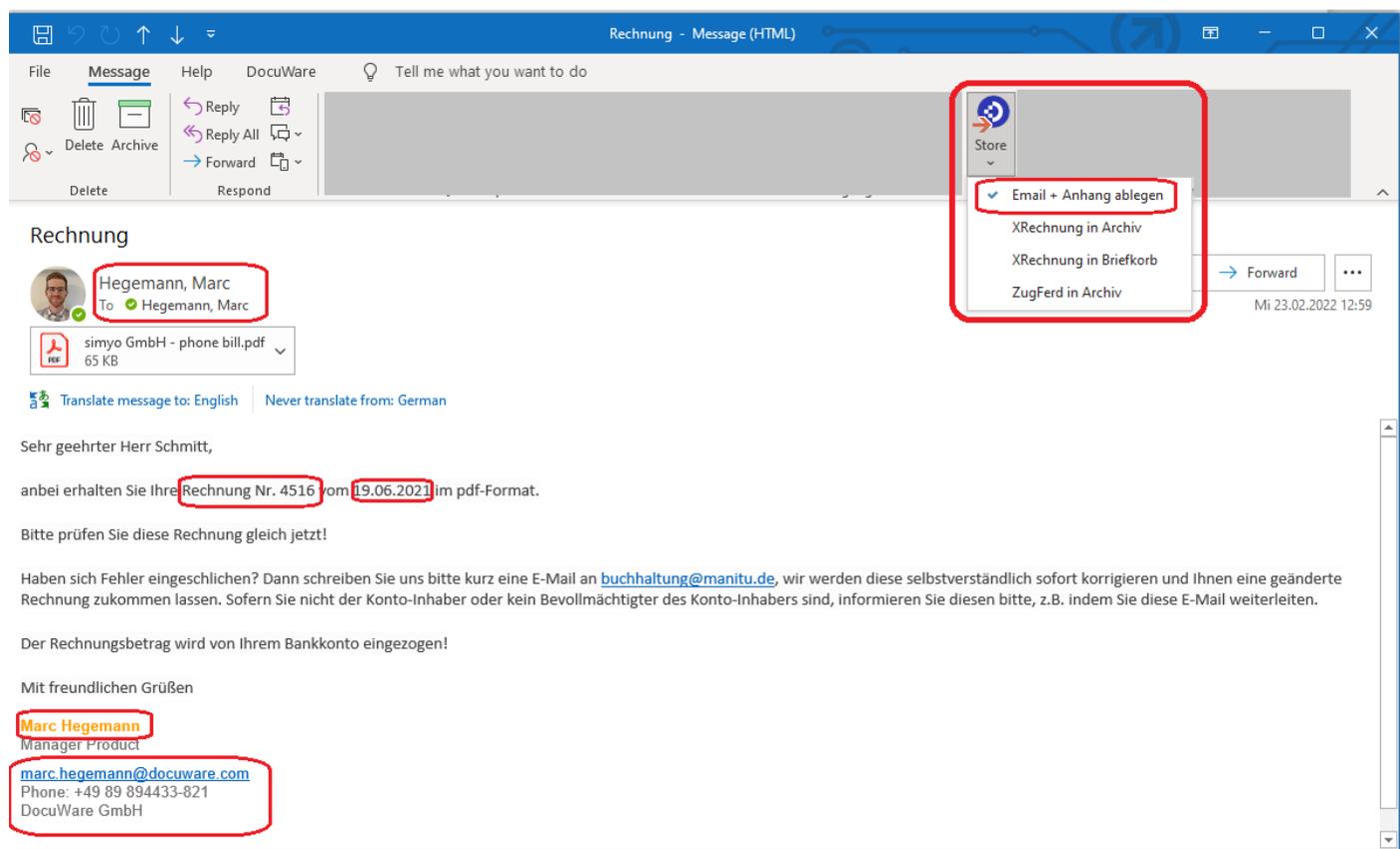
10/06/08

Art.	Best.	Lief.	Artikel	Einzel	Gesamt	Mngt.
4358752	1	1	SERVER 2500/RAID HP 2500 BLACK	5699,00	5699,00	19.0
2394382	1	1	JUREBOX 7000 INCL. 2X WORM DRIVE	25.000,00	25.000,00	19.0

Marc Hegemann  
03.08.2021, 16:34  
Genehmigt

Aufgänger/Other Value	Steuernummer/Invoice No.	Netto/Net Amount	Brutto/Brutto Amount	Steuer/Tax	Summe/Total	Währung/Currency
30.699,00	421,31	0,00	0,00	31.120,31	19,0	5.912,00

# ...aber weniger intelligentes Indexieren von Emails ☹️



## Schlüssel-Informationen

### Absender

- Name der Firma
- Name des Absenders
- Kontaktdaten (Adresse, Email, Telefon)

### Empfänger

- Name des Empfängers

### Allgemein

- Betreff
- Datum

# Anfängliche Hürden

# Roadblocker Compliance

- Dürfen wir Daten unserer Kunden zum Antrainieren einer KI verwenden?
- Zeitaufwändiges Rechtsgutachten mit Analyse der bestehenden Kundenverträge
- Ergebnis: Unter bestimmten Voraussetzungen dürfen Kundendaten verwendet werden.
- Learning: Langfristig sollen unsere Standard-Kundenverträge um einen Passus erweitert werden

# Entwicklung

# Exploration Phase

- Daten Labelling (Enron Datenset)
- AI Modell: NER mit SpaCy

F1 Score*	Sender Adresse	Sender Telefon	Sender Name	Sender Organisation	Empfänger Name
Enron (500 gelabelte Emails)	88%	70%	52%	42%	52%

# Implementation Phase

- Herausforderung Kundendaten
- Neues AI Modell: BiLSTM

Dear colleagues,  
The end of project is coming closer and we would like to reassess the current state of Artificial Intelligence within DocuWare and compare the results.

Best regards,

**Marc Hegemann**  
Manager Product

→ [0, ..., 0, **B-SENDER**, **I-SENDER**, 0, 0]

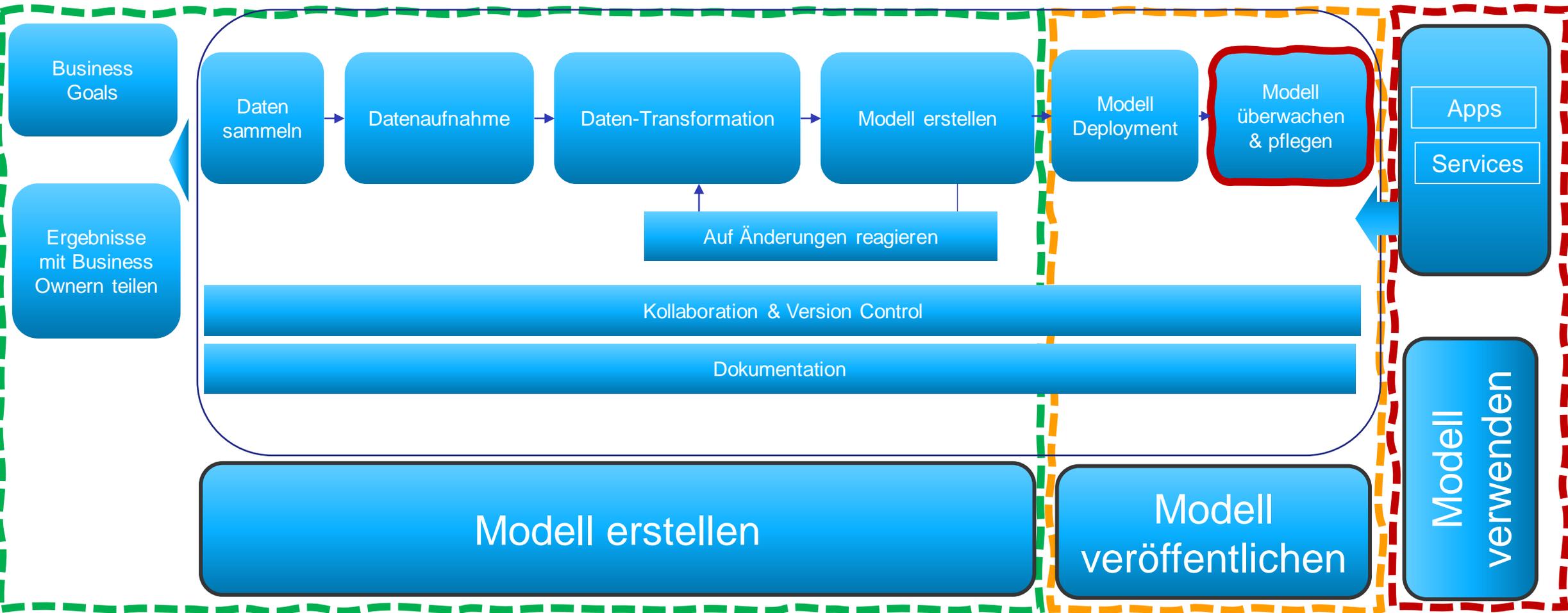
F1 Score	Sender Name	Empfänger Name
Enron	87% (+35%)	71% (+19%)
Kundendaten	78% (---)	?

# Daten & Labelling

- Probleme beim Akquirieren von Daten:
  - Quantitative Unterschiede in den Daten je nach Kunde
  - Für manche Felder keine Label in Kundendaten vorhanden
  - Nicht alle Felder sind in den Emails vorhanden
  - Labeln sehr zeitaufwändig
- Aktuelle Ergebnisse

F1 Score	Sender Name	Empfänger Name	Sender Organisation	Sender Telefon	Sender Adresse
Kunden Label	80% (2.400 Stk.)	Keine Label in Kundendaten	69% (1.200 Stk.)	Kein Label in Kundendaten	Kein Label in Kundendaten
Eigene Label (250 Stk.)	69%	Labelling ongoing	66%	33%	39%

# MLOps + Docuware = Azure Machine Learning



# AI Vision & Learnings

# AI Vision & Company Strategy

**Generate customer value...**

**...through AI technology!**

**Business Decision Preparation**

**Lernende Systeme verstehen das Geschäftsmodell des Anwenders** und helfen täglich bessere Entscheidungen zu treffen

**Business Process Automation**

**DocuWare liefert semantische Einblicke auf Dokumenten- und Prozessebene**, welche dem Anwender eine kontinuierliche Prozessanalyse und -verbesserung erlauben

**Content management**

**KI erweitert die DocuWare Plattform um eine semantische Ebene** welche dem Anwender dabei hilft sein Geschäftsmodell besser zu verstehen und seine Daten effizient und sicher zu archivieren.

# Learnings



Compliance Checks



Data Quality Checks



Data Strategy?



Time Management & Ressourcen



Vielen Dank für Ihr Interesse!

**Ich freue mich auf Ihre Fragen 😊**

Marc Hegemann