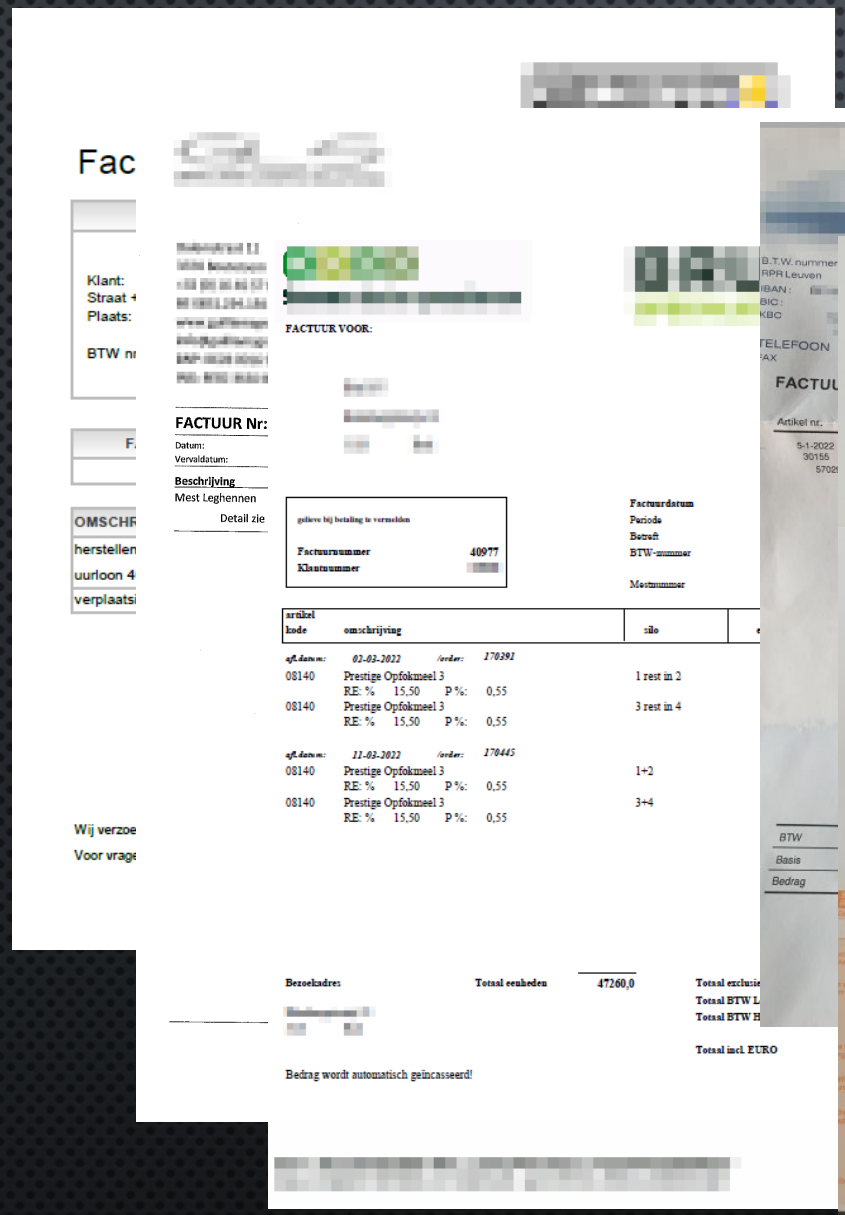
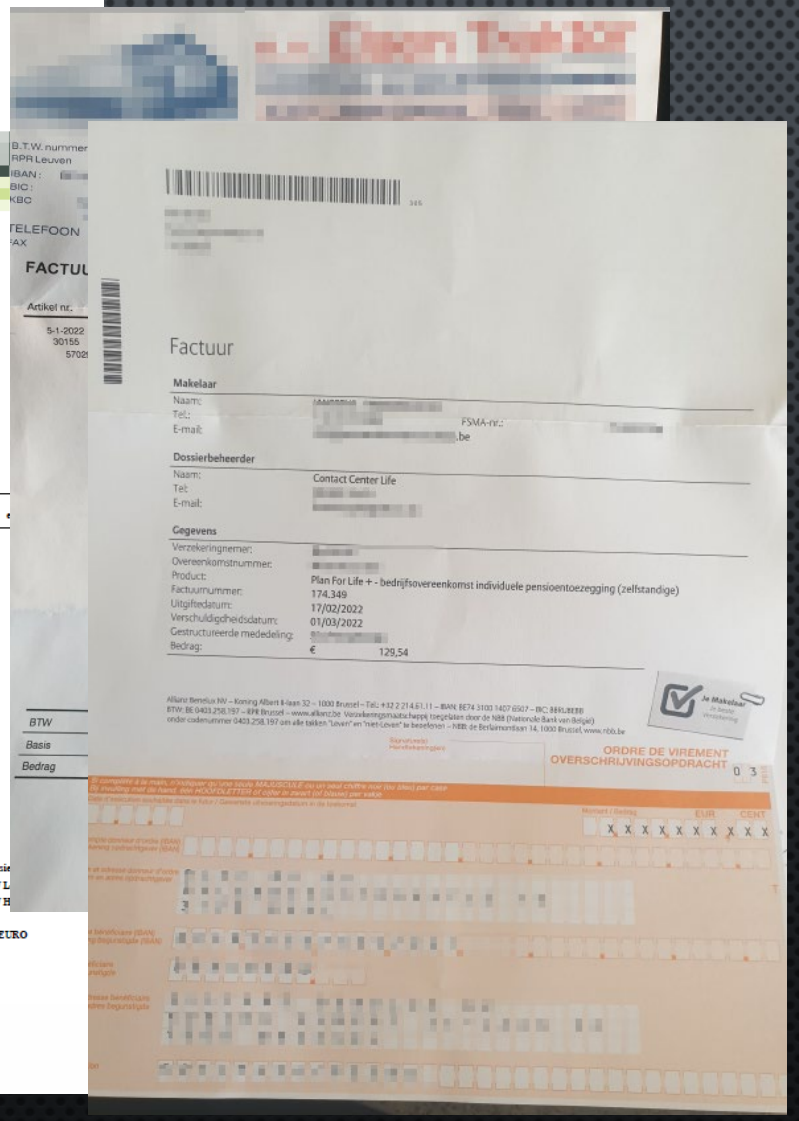
Invoice Processing

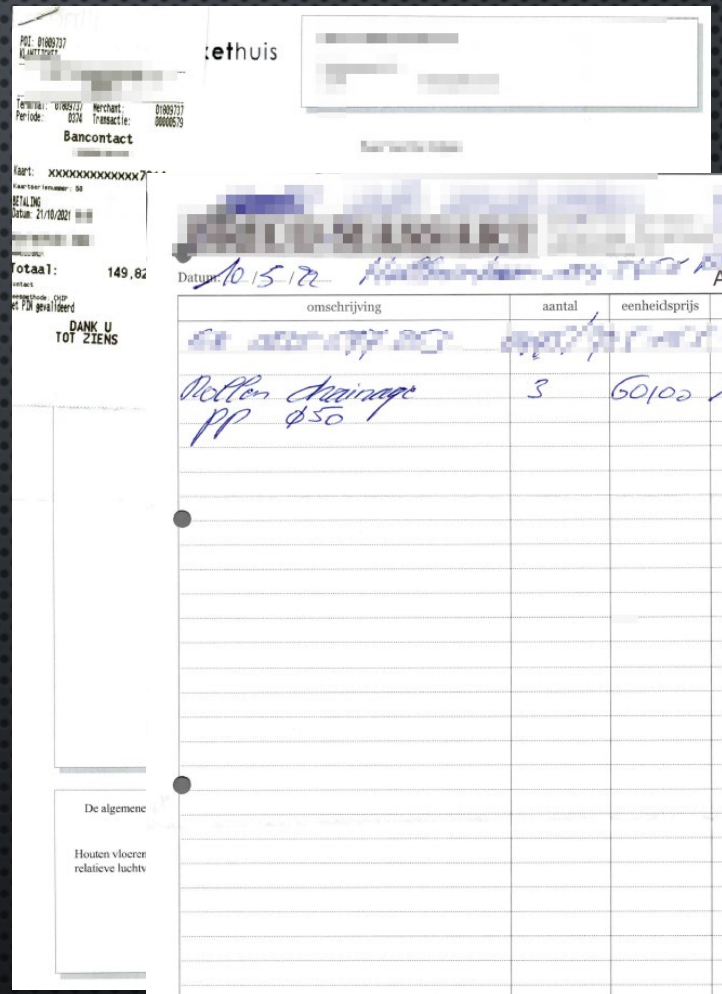
And Leveraging it with AI

By Andy Oeyen

INVOICES


Fac
Klant:
Straat +
Plaats:
BTW nr
FACTUUR VOOR:
FACTUUR Nr:
Datum:
Vervaldatum:
Beschrijving
Mest Leghennen
Detail zie
gelieve bij betaling te vermelden
Factuurnummer 40977
Klantnummer
Factuurdatum
Periode
Betreft
BTW-nummer
Mestnummer
artikel kode
omschrijving
silo
afl.datum: 02-03-2022 /order: 170391
08140 Prestige Opfokmeel 3 1 rest in 2
RE: % 15,50 P %: 0,55
08140 Prestige Opfokmeel 3 3 rest in 4
RE: % 15,50 P %: 0,55
afl.datum: 11-03-2022 /order: 170445
08140 Prestige Opfokmeel 3 1+2
RE: % 15,50 P %: 0,55
08140 Prestige Opfokmeel 3 3+4
RE: % 15,50 P %: 0,55
Bezoekadres:
Totaal eenheden 47260,0
Totaal incl. EURO
Bedrag wordt automatisch geïncasseerd!
Wij verzoe
Voor vrage


B.T.W. nummer
RPR Leuven
IBAN :
BIC :
KBC
TELEFOON
FAX
FACTU
Artikel nr.
5-1-2022
30155
BTW
Basis
Bedrag
Factuur
Makelaar
Naam:
Tel.:
E-mail:
FSMA-nr.:
Dossierbeheerder
Naam:
Tel:
E-mail:
Contact Center Life
Gegevens
Verzekeringnemer:
Overeenkomstnummer:
Product:
Plan For Life + - bedrijfsovereenkomst individuele pensioentoezegging (zelfstandige)
Factuurnummer:
174.349
Uitgiftedatum:
17/02/2022
Verschuldigdheidsdatum:
01/03/2022
Gestructureerde mededeling:
Bedrag:
€ 129,54
ORDRE DE VIREMENT
OVERSCHRIJVINGSOPDRACHT
EUR
CENT


POI: 01989727
ethuis
Bancontact
Totaal: 149,82
DANK U
TOT ZIENS
Datum 10 15 22
A4 07398
omschrijving
aantal
eenheidsprijs
totaal prijs
Rollen draainagel
PP Ø50
3
60,00
180,00
De algemene
Houten vloeren
relatieve luchtv

INVOICE PROCESSING

METHOD 1

MANUAL PROCESSING

Invoice types:
- All kinds

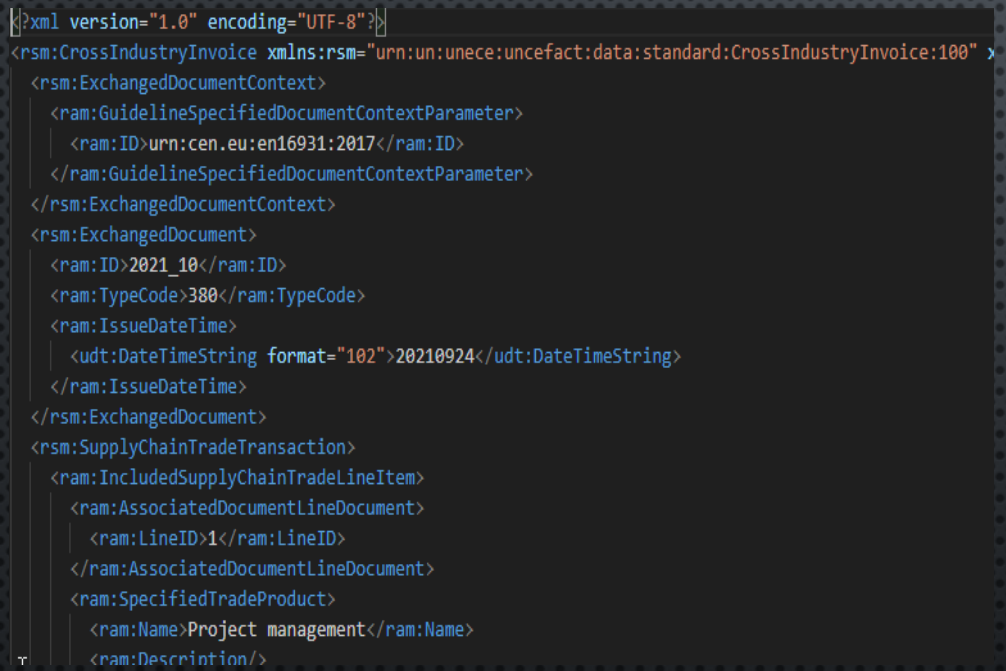
Advantages:
- Trained brains
- Reliable

Disadvantages:
- Slower

INVOICE PROCESSING

METHOD 2

STANDARD E-INVOICES PROCESSING



```xml
<?xml version="1.0" encoding="UTF-8"?>
<rsm:CrossIndustryInvoice xmlns:rsm="urn:un:unece:uncefact:data:standard:CrossIndustryInvoice:100" >
  <rsm:ExchangedDocumentContext>
    <ram:GuidelineSpecifiedDocumentContextParameter>
      <ram:ID>urn:cen.eu:en16931:2017</ram:ID>
    </ram:GuidelineSpecifiedDocumentContextParameter>
  </rsm:ExchangedDocumentContext>
  <rsm:ExchangedDocument>
    <ram:ID>2021_10</ram:ID>
    <ram:TypeCode>380</ram:TypeCode>
    <ram:IssueDateTime>
      <udt:DateTimeString format="102">20210924</udt:DateTimeString>
    </ram:IssueDateTime>
  </rsm:ExchangedDocument>
  <rsm:SupplyChainTradeTransaction>
    <ram:IncludedSupplyChainTradeLineItem>
      <ram:AssociatedDocumentLineDocument>
        <ram:LineID>1</ram:LineID>
      </ram:AssociatedDocumentLineDocument>
      <ram:SpecifiedTradeProduct>
        <ram:Name>Project management</ram:Name>
        <ram:Description/>
```

Type of invoices:

Digital machine readable invoices with standard format (UBL/ CEFACT/ Zugferd/…)

Method:

Generic algorithm

Advantages:

Very (very) high reliability

Fast

Can be combined with specific algorithms

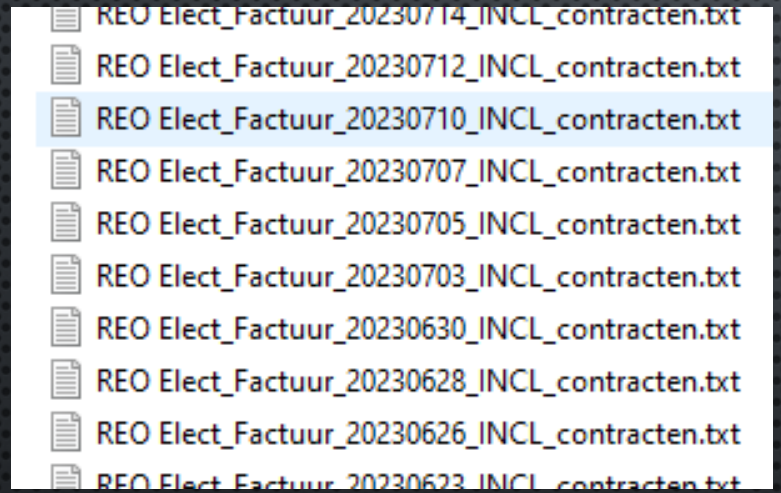
Disadvantage:

Not commonly used in agriculture (in Belgium)

INVOICE PROCESSING

METHOD 3

DIGITAL INVOICE PROCESSING



Type of invoices:
- Digital machine readable invoices
- XML/JSON/CSV/…

Method:
- Specific algorithms

Advantages:
- Very high reliability (depending on how data was generated)
- Fast

Disadvantage:
- Different algorithm per type

INVOICE PROCESSING

METHOD 4

UNSTRUCTURED INVOICE PROCESSING

Type of invoices:
- Digital or Digitalised invoice (unstructured)
- Paper -> pdf, png, …

Method:
- OCR in combination with specific algorithms and templates

Advantages:
- Fast

Disadvantage:
- Result depends on quality of input
- Different algorithm per invoice type
- Changes

INVOICE PROCESSING

METHOD 5

AI INVOICE PROCESSING

Type of invoices:
- Digital or digitalised invoice (unstructured)
- Paper -> pdf, png, …

Method:
- AI model

Advantages:
- Fast
- All kinds of invoice types

Disadvantage:
- Accuracy depends on model
- Black box is scary
- Training takes time

Side note

- Most invoice processing methods Still require some kind of mapping to a standard product / product type.

AI PROCESSING: STEP 1

AI PROCESSING: STEP 2

Label invoices with
low probabilities (Active learning) or known mistakes

Proof of concept AI implementation

- Model Microsoft LayoutLMv3
  - Natural language processing
  - Combines text with layout
  - Tesseract for OCR
  - Usage: Finetuned the base model

- Label software: Label-studio
  - Connects to ML-AI backend
  - Active learning possible

Step 1:
Initial prediction with
Base model

->All Labels: OTHER (red)

Step 2:
Labels invoices

Step 3:
Finetune model

Step 4:
Prediction with finetuned model

Results and thoughts

- F1 after 1st iteration of training with 56 labelled invoices : 0,82
- Further finetuning is necessary
- Preprocessing model to improve overall accuracy
  - Handwritten invoices/ receipts /digital invoices/scanned
- Standard e-invoices would make this obsolete
- Other ideas are welcome
- Thank you for your attention

MORE INFORMATION:

- https://github.com/microsoft/unilm/blob/master/layoutlmv3/README.md
- https://labelstud.io/