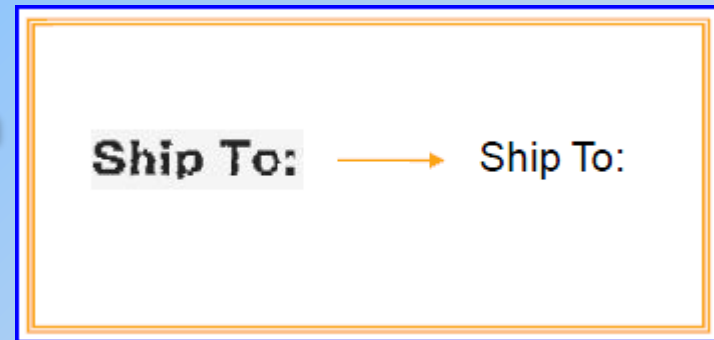


OCR Technieken

- OCR – Optical Character Recognition
 - ICR – Intelligent Character Recognition
 - OMR – Optical Mark Recognition
 - Barcode
 - Handschrift
-
- Ondersteuning image processing
 - Vaste formulier verwerking
 - Semi-gestructureerde formulier verwerking
 - Ongestructureerde Dokument verwerking

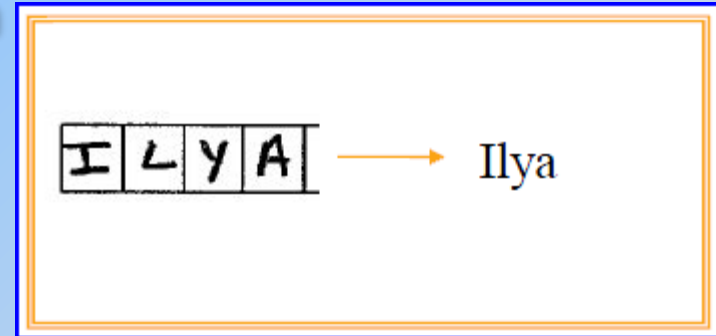
OCR Technieken

- OCR – Optical Character Recognition
 - ICR – Intelligent Character Recognition
 - OMR – Optical Mark Recognition
 - Barcode
 - Handschrift
-
- Ondersteuning image processing
 - Vaste formulier verwerking
 - Semi-gestructureerde formulier verwerking
 - Ongestructureerde Dokument verwerking



OCR Technieken

- OCR – Optical Character Recognition
- ICR – Intelligent Character Recognition
- OMR – Optical Mark Recognition
- Barcode
- Handschrift



- Ondersteuning image processing
- Vaste formulier verwerking
- Semi-gestructureerde formulier verwerking
- Ongestructureerde Dokument verwerking

OCR Technieken

- OCR – Optical Character Recognition
- ICR – Intelligent Character Recognition
- OMR – Optical Mark Recognition
- Barcode
- Handschrift

The diagram illustrates an OMR form with two rows of options. The first row is 'Settlement Account' with an unchecked checkbox. The second row is 'Card Account' with a checked checkbox. An arrow points from the checked checkbox to the text 'Card Account' on the right side of the form.

Settlement Account	<input type="checkbox"/>	→	Card Account
Card Account	<input checked="" type="checkbox"/>		

- Ondersteuning image processing
- Vaste formulier verwerking
- Semi-gestructureerde formulier verwerking
- Ongestructureerde Document verwerking

OCR Technieken

- OCR – Optical Character Recognition
 - ICR – Intelligent Character Recognition
 - OMR – Optical Mark Recognition
 - **Barcode**
 - Handschrift
-
- Ondersteuning image processing
 - Vaste formulier verwerking
 - Semi-gestructureerde formulier verwerking
 - Ongestructureerde Dokument verwerking



OCR Technieken

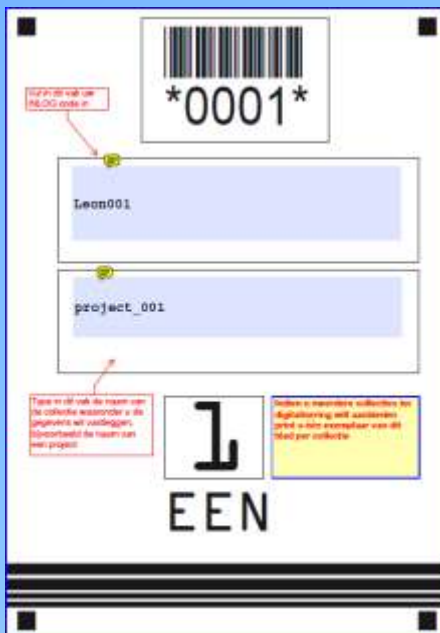
- OCR – Optical Character Recognition
 - ICR – Intelligent Character Recognition
 - OMR – Optical Mark Recognition
 - Barcode
 - Handschrift
-
- Ondersteuning image processing
 - Vaste formulier verwerking
 - Semi-gestructureerde formulier verwerking
 - Ongestructureerde Dokument verwerking



* CRITICAL* → * Critical *

Ondersteuning Image processing

- Gebruik van OCR techniek om archiefstructuren vast te leggen
 - Bijvoorbeeld door barcodes, zone herkenning of combinatie van factoren.



Waste Formulier Verwerking

FORM 01 CUSTOMER'S REFERENCE DATA CONFIRMATION

Mark

PERSONAL DETAILS

Mr Mrs Ms Miss

First Name: ILYA Middle Initial:

Last Name: EVDOKIMOV Nationality: RUSSIAN

City of Birth: TASHKENT

Birth Date (mm/dd/yy): 12/21/2982 Social Security Number: 123456789

Marital Status: Married Single Separated/Divorced Other Dating

Spouse First Name: Middle Initial: Spouse Last Name: Mother's Maiden Name: IYAMOVA

ADDRESS/CONTACT DETAILS

HOME ADDRESS

Country: USA ZIP Code: 94538

State: CA City: FREEMONT

Street: 47221 FREEMONT BLVD

Home Phone: 510-226-6717 Mobile Phone:

PRESENT ADDRESS

Country: ZIP Code:

State: City:

Street:

Home Phone: Building: Apartment:

PROFESSIONAL DETAILS

Employed Self-employed Retired Other

Name of Employer or Self-Employed Company Name: FLSH AND CHEPS CO.

Occupation: RAW FISH TASTING MANAGER

Country: USA ZIP Code: 94538

State: CA City: FREEMONT

Street: 47221 FREEMONT BLVD

Home Phone: Building: Apartment:

Work Phone: 510-226-6717 Mobile Phone:

Gross Annual Income: 15000 USD Employment Date (mm/yy): 01/01/1999

I Declare the Correctness of the Information above

Document Jels (mm/dd/yy): 01/06/2006

Stamp:

0122095161180

→ Name: Ilya
Date: 12/21/2982

Semi gestructureerde formulieren

206185

Invoice

AA APPAREL
1 Memphis Dr
Memphis, TN 38114
128-450-7890

RECEIVED MAR - 9

DATE: 06/09/04 INVOICE NO: 99044

SHIP TO: US NORTHLAND
Doug Daniel
18 Prince George Drive
Prince George, VA 23875

P.O. NO.	TERMS	REP	SHIP DATE	SHIP VIA	FOS	PROJECT
1505	Net 30		06/09/04	Country		
QTY	ITEM	DESCRIPTION		RATE	AMOUNT	
7	4990	Gloves or mittens, NCI, in boxes		150.00	1,050.00	
42	15060	Bulk, NCI, inflated		275.00	11,550.00	
24	15520	Atletic or Sporting Goods, NCI		317.00	7,608.00	
TOTAL					20,208.00	

INVOICE NUMBER
24567

DATE: 06/09/04 CUSTOMER NUMBER: 35594

SHIP TO: CUTLER - WARDEN CANADA
150 HANCOCK STREET
TOMAWANDA NY 14180

SERVICE NO.	QUANTITY	UNIT PRICE	TAX	PRODUCT TOTAL
000000004	18.00		50.00	
	13.00		138.00	
	7.00		76.00	
	12.00		66.00	
	8.00		36.00	
	6.00		25.00	
	11.00		113.00	
	10.00		66.00	
	2.00		76.00	
	5.00		50.00	
	3.00		15.00	
	16.00		183.00	
	7.00		35.00	
	8.00		36.00	
	18.00		100.00	
	11.00		33.00	
	2.00		20.00	
	14.00		180.00	
	10.00		180.00	
	5.00		25.00	
	7.00		24.00	
	2.00		24.00	
	5.00		53.00	
	10.00		-19.00	
	3.00		33.00	
	3.00		-88.00	
	50.00		23.00	
	11.00		25.00	
	7.00		31.00	
	11.00		176.00	
	13.00		130.00	
	5.00		25.00	

Invoice No: 99044
Date: 06/09/04

→ Invoice No: 24567
Date: 06/09/04 (06/09/2004)

Ongestructureerde documenten



Vastleggen:

- Datum
- Afzender
- Contactpersoon

Techniek, de standaard mogelijkheden

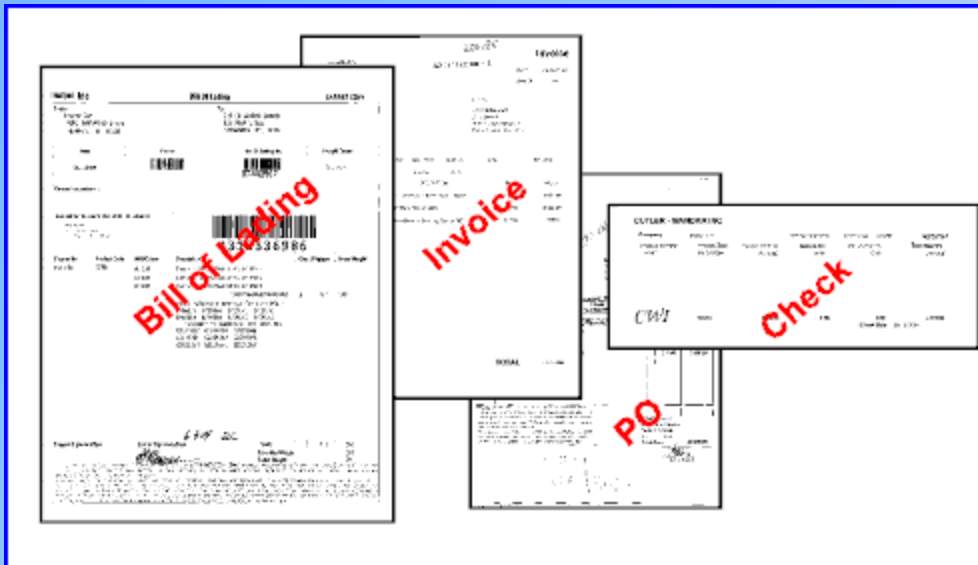
- Volledige tekst herkenning
- Classificeren
- Index velden uitlichten
- Redactioneel (b.v. anonimiseren van correspondentie)
- Routeren van poststukken
- Automatisch archiveren
- Hergebruik van teksten
- Image Rotatie

Techniek: volledige tekst herkenning

- Image bestand naar elektronisch data bestand
- ALLE text op de pagina
- Inclusief:
 - Image Pre-processing
 - Document Analyse/Zone herkenning
 - Data extractie
 - Export (momenteel PDF, DOC)

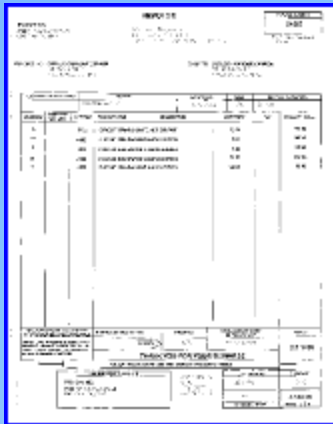
Techniek: Classificeren

- De Software herkent de dokument soorten
- Scan batches van gemengde documenten



Techniek: Indexvelden uitlichten

- Alleen de gewenste velden selecteren
- Normaliseren van data
- Export normaal gesproken naar een database



The screenshot shows a software interface with a data table. The table has several columns, including 'Factuurdatum', 'Afzender', 'Totaal bedrag', and 'BTW bedrag'. The data is organized into rows, with some rows highlighted in blue. The interface also includes various form fields and buttons, typical of a data management application.

- Factuurdatum
- Afzender
- Totaal bedrag
- BTW bedrag

Techniek: Hoe accuraat

- Of liever, hoe bepaal je de accuratesse
 - Dokumentsoort accuratesse
 - Veld/Zone locatie accuratesse
 - Datasoort accuratesse
 - Woord accuratesse
 - Karakter/cijfer accuratesse
- Bedenk dat fouten niet altijd zichtbaar zijn en bepaal daarom van tevoren de invloed van mogelijke herkenningsfouten.

Techniek: standaard toepassingen

- Dokument Conversie
- Dokument Archiveren / Terugzoeken
- Faktuur verwerking
- Verzekering verwerking(medisch, hypotheken)
- Vrachtbrief verwerking
- Enquete verwerking

Wie zijn de belangrijkste producenten

- Er zijn feitelijk maar 4 producenten van OCR technologie
 - Heel veel producten gebruiken dus de kerntechnologie van deze 4.

Wie zijn de belangrijkste producenten

- ABBYY
- Nuance (voorheen ScanSoft)
- ReadI.R.I.S
- Océ

- CharacTell
- ParaScript
- A2iA

- Handvol Open Source
- Handvol andere verkooporganisaties

Wie gebruiken deze engines

- AnaComp
- Anydoc
- BancTec
- BrainWare
- Captaris
- Captivation
- Cardiff
- Cvision
- DataCap
- DigiTech
- eCopy
- EMC Documentum
- Kofax
- LaserFiche
- LeadTools
- Microsoft
- NSi AutoStore
- OnBase
- Perceptive Imaging
- ReadSoft
- SER
- Top Image Systems
- Tower
- Westbrook
- Xerox
- Tientallen anderen

Koper: pas op

- 30% van de organisaties die OCR pakketten kopen, kiezen het verkeerde pakket,
- Meer dan 50 % van de organisaties die OCR pakketten kopen gebruiken de software verkeerd, of niet volledig.

Koper: pas op

- Als OCR de achterliggende reden is voor een aankoop beslissing, weet dan welke engine erachter zit!
- Bespreek en beoordeel de totale oplossing en niet slechts de onderdelen!
- Kijk voorbij de “gimmicks”
- Zorg dat u zeker kunt zijn dat uw verkoper/producent uw vertrouwen waard is.

Koper: pas op, welke engine?

- Welke versie van de engine is gebruikt.
- Wat is de upgrade strategie van de leverancier ten aanzien van updates van de kern engine.

Koper: Pas op, bespreek het totaal

- Scanner / Input
- Index gegevens
- Opslag
- Zorg tevoren voor een lijst met essenties.

Koper: Pas op, kijk door “gimmicks”

- NIETS! Is 100%
- Alle voorbereide demo's werken perfect
- Test het systeem altijd met eigen dokumenten
- Versie nummers kunnen betrekking hebben op iets anders.

Koper: Pas op, vertrouwen

- Support na installatie (test het)
- Waar krijgt u professionele service
- Begrijpt de leverancier het totaalbeeld en niet alleen maar de onderdelen?

Bron: Chris Riley, ecm^p

- <http://www.livinganalytics.com/index.htm>
- Vertaald door: Leon van Oosterom, cdia+
 - <http://www.elveo.nl>