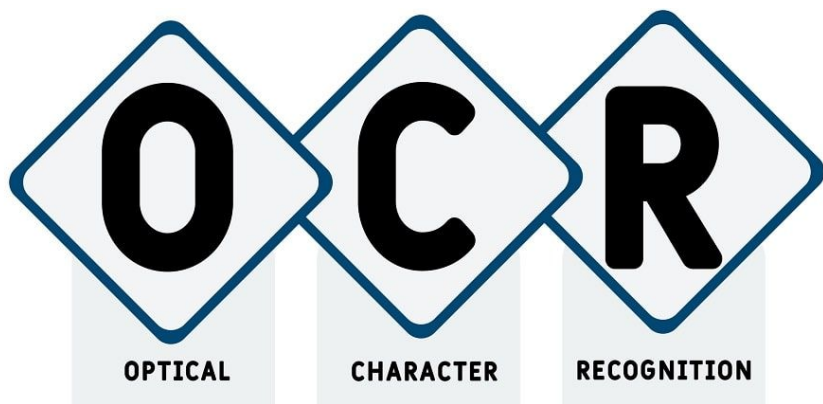


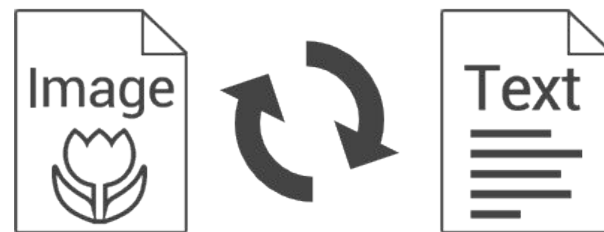
## A collection of business-related icons. The central focus is a magnifying glass with a blue star inside it, resting on a document. To the left is a checklist with three items, the first two marked with checkmarks and the last with an 'X', and a blue pen. Above the magnifying glass is a target with an arrow in the bullseye. To the right is a document with a person icon. Below the magnifying glass is an envelope icon. In the bottom left is a megaphone. The background is filled with various faint, repeating icons such as gears, sliders, and document outlines.

# SOMMAIRE

1. Définition et application
2. Données et méthodes utilisées
3. Conclusion



La reconnaissance optique des caractères, ou *Optical Character Recognition* est une conversion électronique d'images textuelles dactylographiées, manuscrites ou imprimées. Ce texte est encodé pour l'obtenir en format texte.



**Domaine d'application où l'OCR est le plus utilisée :**  
Domaine bancaire, assurances et valeurs mobilières

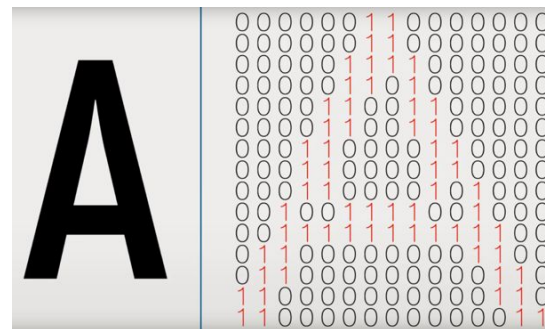
- Optimiser gestion documentaire d'une entreprise
- Détection de documents falsifiés
- Récupération d'information

## DÉFINITION

## Fonctionnement de l'OCR

L'algorithme de détection définit un caractère en évaluant ses lignes et ses traits.

La reconnaissance des motifs, fonctionne en identifiant l'ensemble du caractère.



Reconnaissance des motifs sur un caractère



Détection des lignes et traits



Reconnaissance des motifs sur une ligne de texte.

## DÉFINITION

## Fichiers format .pdf

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
145	6.7	3.3	5.7	2.5	virginica
146	6.7	3.0	5.2	2.3	virginica
147	6.3	2.5	5.0	1.9	virginica
148	6.5	3.0	5.2	2.0	virginica
149	6.2	3.4	5.4	2.3	virginica
150	5.9	3.0	5.1	1.8	virginica

**Inspection and Surveillance Clearance Note**

Type	Frequency	Comments	Notes	Responsibility
How inspection frequency target time achieved for the batch(es) concerned?	Y	Except for activities that haven't occurred yet.	<p>Note: This refers to the completion of the surveillance of the batch(es) at the time of release of the year.</p> <p>2) Where FQMC, it is important to indicate this inspection meeting will be required (with mg, DPM and DPM) - proposed action possible in addition of the Clearance Note (to be done parallel with the issuance of the Clearance Note)</p>	Activity Plant/Regimes Leader
If inspection target <u>has not</u> been achieved, indicate what the inspection gaps are (missing inspections) are.	N		For example, provide list of the guides that have not been implemented but should have been.	Activity Plant/Regimes Leader
If inspection target <u>has not</u> been achieved, provide justification why it is unacceptable for the batch(es) to be released.	Y		For example, provide argument that the guides not applied have no direct link with the quality of the products. In that guide not applied do not jeopardize the capability of the supplier, or the related operations due to the knowledge of the result of the past 10 surveillance DPM in particular). All request that could give a level of confidence on the quality of the product for the operations for which several guides applied.	Activity Plant/Regimes Leader
No. of Observation Sheets issued		N/A	N/A	Activity Plant/Regimes Leader
How all observation sheets have collectively responded to and closed at time of issuing the clearance note?		N/A	<p>2) Provide list of open observation sheets.</p> <p>2) Justify to decide why the Observation Sheets remain open.</p> <p>If indicate whether sheet has been closed to the Project Manager.</p> <p>If the sheet has remained with open Observation Sheet, indicate why.</p>	Activity Plant/Regimes Leader
Should a final inspection be targeted before the OMC is accepted? (proposal of 1 or 0) (to be indicated by HPC)	N		<p>If targeted, a final report will be required within 48 hours of completion of inspection, in addition to Surveillance Action Report.</p> <p>Important! Note: HPC environment of the final inspection is required.</p>	Equipment Engineers/Engineers Use Plant

**Completion Contact Details**

Name	Phone	Date
Activity Plant AP (for OMC)		
Project Leader - AL (for OMC)	5.19622	Feb 22/1 2022
Production Engineer - AS (for OMC)		
Inspection Plant - AP (for OMC)		

**MANUFACTURING SURVEILLANCE CLEARANCE NOTE ISSUED BY DIRECTION INDUSTRIELLE TO HPC**

1. **Contract Details**

HPC Contract Number	006320
Batch / Package to be released to HPC (including Quality Reference Certificate)	Bundle B-002 146
Supplier Contract	1801

2. **Details of Surveillance Offer**

Manufacturing Surveillance Offer Reference	2020L332038801 and 8
DI/IRA Manufacturing Surveillance Unit	DI 512

3. **Status of In-Factory Inspections**

**Completion of Inspection Guides**

Surveillance Guide is scheduled in Surveillance offer	AS12	MRP01	MS12	MS14	MS16	MS20	MS24
Additional Guides programmed by Inspection in WCC/Task/DOC			1				
Name of Guide	Working Procedure	Start-Up Manufacturing	Pre-requisites release	Examination of sub-welds	Sub-suppliers surveillance	Base endurance control	Final Treatment
Scheduled frequency (as in Surveillance Offer)	1	1	1	1	1	1	1 per HT type
Actual frequency completed	1	1	1	1	1	1	1

LES DONNÉES

- **Tabula**: Outil utilisé pour la lecture de tableaux de données dans les PDFs. Tabula est très utilisé par les sociétés d'informations.

The New York Times



- **pdfPlumber**: Package open-source de Python. Outil simple d'utilisation, et très efficace pour la récupération de texte et données dans les tableaux d'un PDF.
- **Tesseract**: Logiciel de reconnaissance optique de caractère. Peut lire des caractères de type UTF-8 dans plus de 100 langues.
- **OpenCV**: OK bon app

LES MÉTHODES / LIBRAIRIES UTILISÉES

**Fresh Bakery**  
3494 Rosewood Court  
ELM MOTT  
Texas

**INVOICE**  
# 6

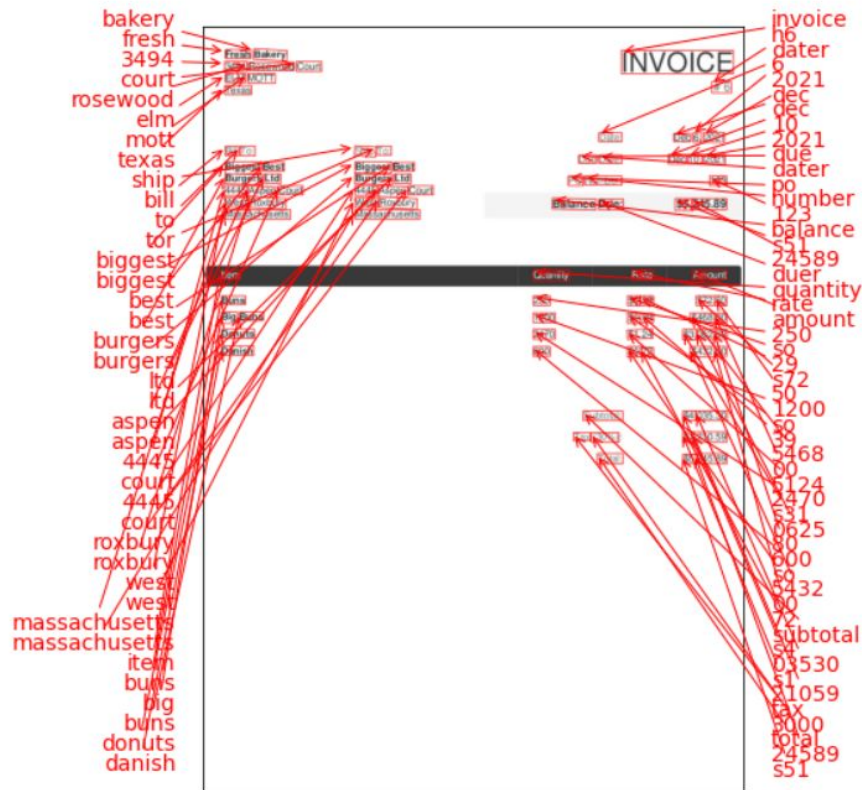
Bill To:  
**Biggest Best Burgers Ltd**  
4445 Aspen Court  
West Roxbury  
Massachusetts

Ship To:  
**Biggest Best Burgers Ltd**  
4445 Aspen Court  
West Roxbury  
Massachusetts

Date: Dec 6, 2021  
Due Date: Dec 10, 2021  
PO Number: 123

**Balance Due: \$5,245.89**

Item	Quantity	Rate	Amount
Buns	250	\$0.29	\$72.50
Big Buns	1200	\$0.39	\$468.00
Donuts	2470	\$1.24	\$3,062.80
Danish	600	\$0.72	\$432.00
		Subtotal:	\$4,035.30
		Tax (30%):	\$1,210.59
		Total:	\$5,245.89



Fresh Bakery

3494 Rosewood Court

ELM MOTT

Texas

INVOICE

# 6

Bill To:

Biggest Best

Burgers Ltd

4445 Aspen Court

West Roxbury

Massachusetts

Ship To:

Biggest Best

Burgers Ltd

4445 Aspen Court

West Roxbury

Massachusetts

Date:

Dec 6, 2021

Due Date:

Dec 10, 2021

PO Number:

123

Balance Due:

\$5,245.89

Item	Quantity	Rate	Amount
Buns	250	\$0.29	\$72.50
Big Buns	1200	\$0.39	\$468.00
Donuts	2470	\$1.24	\$3,062.80
Danish	600	\$0.72	\$432.00
Subtotal:			\$4,035.30
Tax (30%):			\$1,210.59
Total:			\$5,245.89

Fresh Bakery

3494 Rosewood Court

ELM MOTT

Texas

INVOICE

# 6

Bill To:

Biggest Best

Burgers Ltd

4445 Aspen Court

West Roxbury

Massachusetts

Ship To:

Biggest Best

Burgers Ltd

4445 Aspen Court

West Roxbury

Massachusetts

Date:

Dec 6, 2021

Due Date:

Dec 10, 2021

PO Number:

123

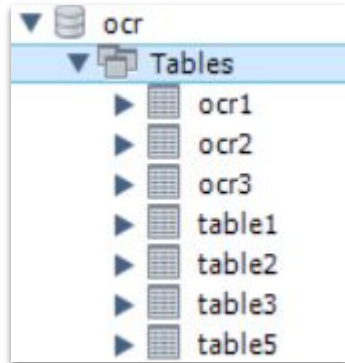
Balance Due:

\$5,245.89

Item	Quantity	Rate	Amount
Buns	250	\$0.29	\$72.50
Big Buns	1200	\$0.39	\$468.00
Donuts	2470	\$1.24	\$3,062.80
Danish	600	\$0.72	\$432.00
Subtotal:			\$4,035.30
Tax (30%):			\$1,210.59
Total:			\$5,245.89

TESSERACT GOOGLE

- Export des données dans une base de données relationnelle via Python avec la librairie SQL Alchemy



A screenshot of a database management tool's table list. The 'ocr' database is expanded, showing a folder named 'Tables' which contains several tables: ocr1, ocr2, ocr3, table1, table2, table3, and table5. Each table is represented by a small grid icon.

ocr
Tables
ocr1
ocr2
ocr3
table1
table2
table3
table5

Unnamed: 0	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21	6	160	110	3.9	2.62	16.46	0	1	4	4
Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.46	20.22	1	0	3	1
Duster 360	14.3	8	360	245	3.21	3.57	15.84	0	0	3	4

- Pour pallier aux erreurs d'exportation vers MYSQL notamment à cause de titre trop long dans nos tableaux, nous avons effectué une modification des données en Python.

Tesseract Google

## OpenCV

PDFPLumber

## Tabula

## Comprendre l'OCR, le guide



## SOURCES