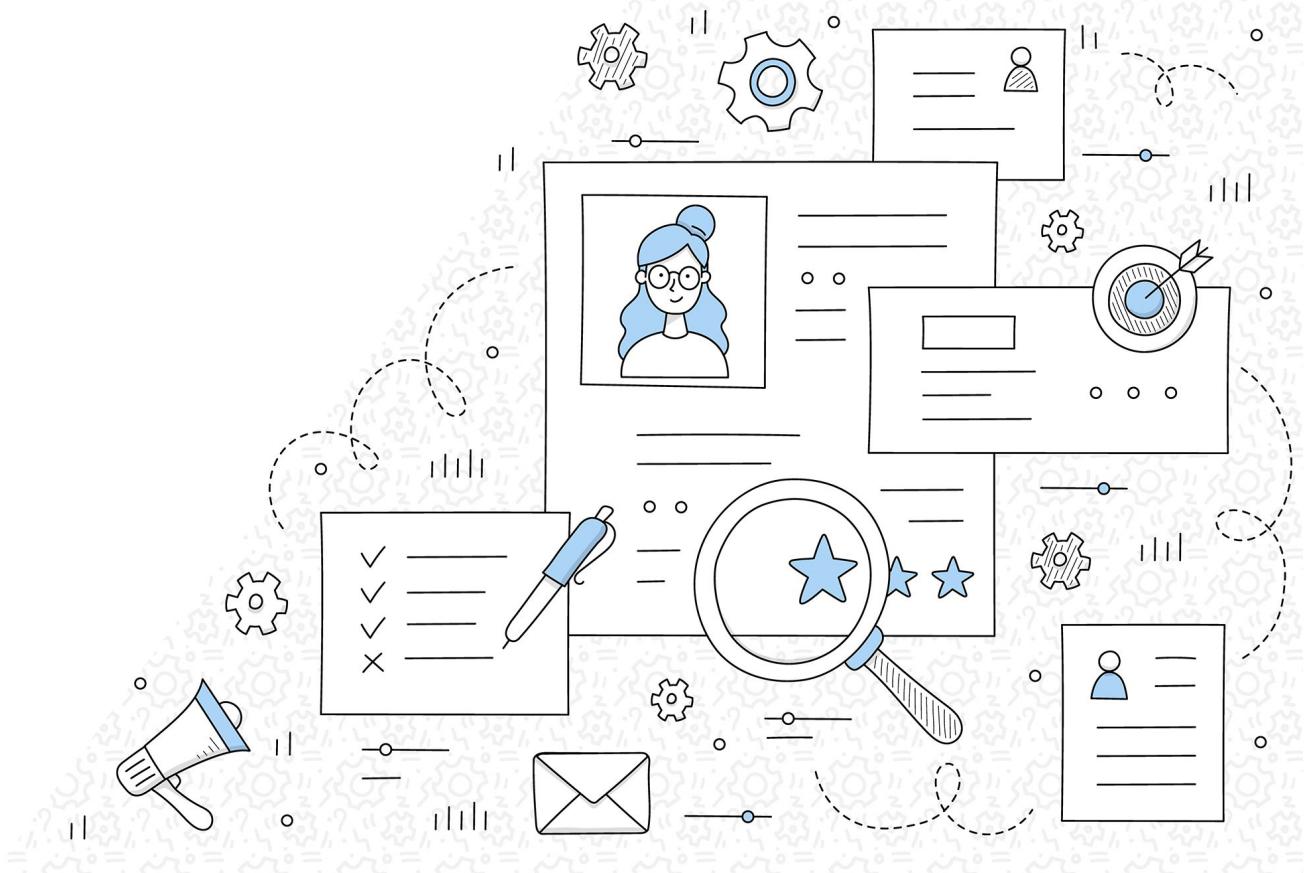
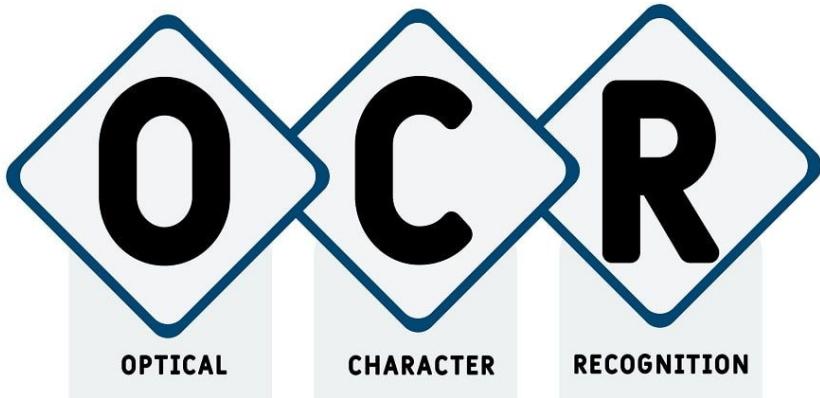


OCR

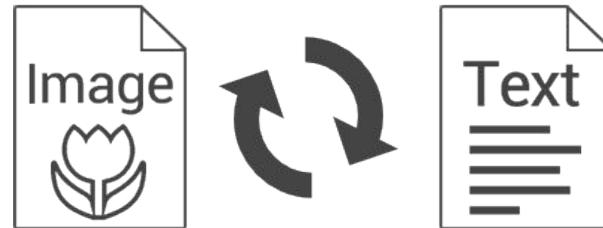


SOMMAIRE

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.



Détection des lignes et traits



Reconnaissance des motifs sur un caractère



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

Manufacturing Surveillance Clearance Note issued by Direction Industrielle to HPC																																														
1 Contract Details																																														
<table border="1"> <tr> <td>HPC Contract Number</td> <td>000000</td> </tr> <tr> <td colspan="2">Batch / Package is for release related to HPC implying Quality Release Certificate</td> </tr> <tr> <td colspan="2">Batch ID: B-000-000</td> </tr> <tr> <td colspan="2">Release Commitment: 100%</td> </tr> </table>					HPC Contract Number	000000	Batch / Package is for release related to HPC implying Quality Release Certificate		Batch ID: B-000-000		Release Commitment: 100%																																			
HPC Contract Number	000000																																													
Batch / Package is for release related to HPC implying Quality Release Certificate																																														
Batch ID: B-000-000																																														
Release Commitment: 100%																																														
2 Details of Surveillance Offer																																														
<table border="1"> <tr> <td>Manufacturing Surveillance Officer Reference</td> <td>DIRECTION INDUSTRIELLE</td> </tr> <tr> <td>Manufacturing Surveillance Officer</td> <td>DI SLI</td> </tr> </table>					Manufacturing Surveillance Officer Reference	DIRECTION INDUSTRIELLE	Manufacturing Surveillance Officer	DI SLI																																						
Manufacturing Surveillance Officer Reference	DIRECTION INDUSTRIELLE																																													
Manufacturing Surveillance Officer	DI SLI																																													
3 Status of In-factory Inspections																																														
<table border="1"> <tr> <td colspan="7">Completion of Inspection Guides</td> </tr> <tr> <td>Surveillance Guides Scheduled in Surveillance Offer</td> <td>0001</td> <td>000002</td> <td>0002</td> <td>0010</td> <td>0016</td> <td>0020</td> </tr> <tr> <td colspan="7">Additional Guides programmed by Inspector in W/C Task/FAC</td> </tr> <tr> <td>Name of Guide</td> <td>Working procedure</td> <td>Exit-CR Manufacturing</td> <td>Pre-inspection release</td> <td>Notification of sub-orders</td> <td>Sub-supplier surveillance</td> <td>New evidence control</td> </tr> <tr> <td>Scheduled Frequency (as in Surveillance Offer)</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0 per HPC type</td> </tr> <tr> <td>Actual Frequency completed</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> </tr> </table>					Completion of Inspection Guides							Surveillance Guides Scheduled in Surveillance Offer	0001	000002	0002	0010	0016	0020	Additional Guides programmed by Inspector in W/C Task/FAC							Name of Guide	Working procedure	Exit-CR Manufacturing	Pre-inspection release	Notification of sub-orders	Sub-supplier surveillance	New evidence control	Scheduled Frequency (as in Surveillance Offer)	0	0	0	0	0	0 per HPC type	Actual Frequency completed	0	0	0	0	0	0
Completion of Inspection Guides																																														
Surveillance Guides Scheduled in Surveillance Offer	0001	000002	0002	0010	0016	0020																																								
Additional Guides programmed by Inspector in W/C Task/FAC																																														
Name of Guide	Working procedure	Exit-CR Manufacturing	Pre-inspection release	Notification of sub-orders	Sub-supplier surveillance	New evidence control																																								
Scheduled Frequency (as in Surveillance Offer)	0	0	0	0	0	0 per HPC type																																								
Actual Frequency completed	0	0	0	0	0	0																																								
<table border="1"> <tr> <td colspan="7">Completion Contract Details</td> </tr> <tr> <td>Date</td> <td>Name</td> <td>Date</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Activity Lead - IP (for example)</td> <td>L.PHEDS</td> <td>Feb.22/2022</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Region Leader - RL (for example)</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Equipment Engineer - EE (for example)</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Inspection Lead - IP (for example)</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>					Completion Contract Details							Date	Name	Date					Activity Lead - IP (for example)	L.PHEDS	Feb.22/2022					Region Leader - RL (for example)							Equipment Engineer - EE (for example)							Inspection Lead - IP (for example)						
Completion Contract Details																																														
Date	Name	Date																																												
Activity Lead - IP (for example)	L.PHEDS	Feb.22/2022																																												
Region Leader - RL (for example)																																														
Equipment Engineer - EE (for example)																																														
Inspection Lead - IP (for example)																																														

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

PROPUBLICA

THE 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

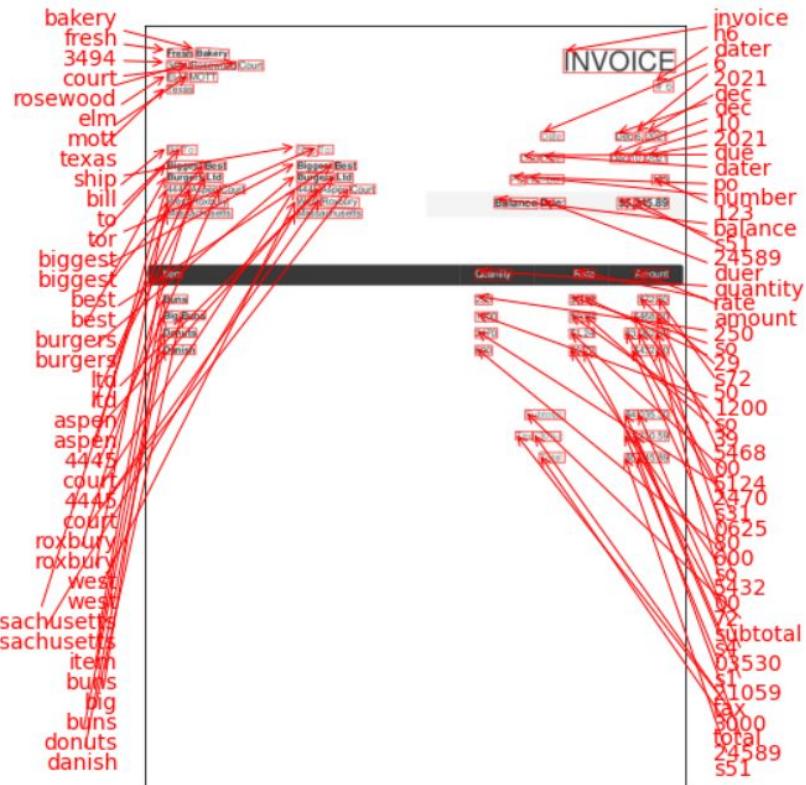
INVOICE

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



KERAS_OCR

Fresh Bakery
3494 Rosewood Court
ELM MOTT
Texas

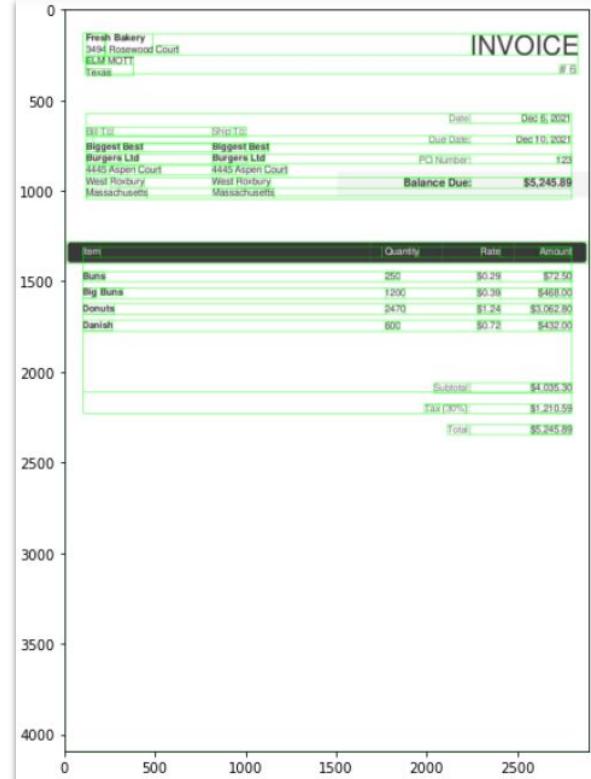
INVOICE

6

Bill To:	Ship To:	Date:	Dec 6, 2021
Biggest Best Burgers Ltd 4445 Aspen Court West Roxbury Massachusetts	Biggest Best Burgers Ltd 4445 Aspen Court West Roxbury Massachusetts	Due Date:	Dec 10, 2021
		PO Number:	123
		Balance Due:	\$5,245.89

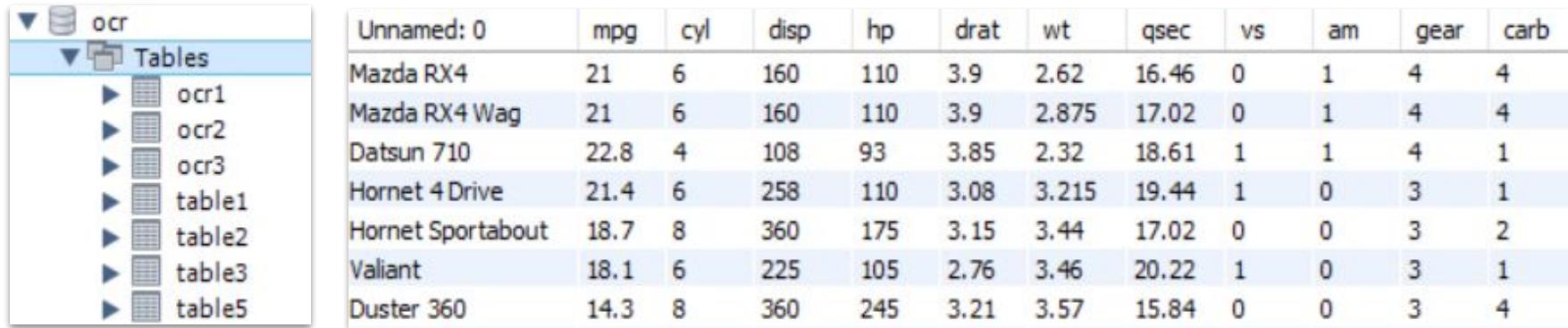
Item	Quantity	Rate	Amount
Buns	250	\$0.29	\$72.50
Big Buns	1200	\$0.39	\$468.00
Donuts	2470	\$1.24	\$3,062.80
Danish	600	\$0.72	\$432.00

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



	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.

MySQL

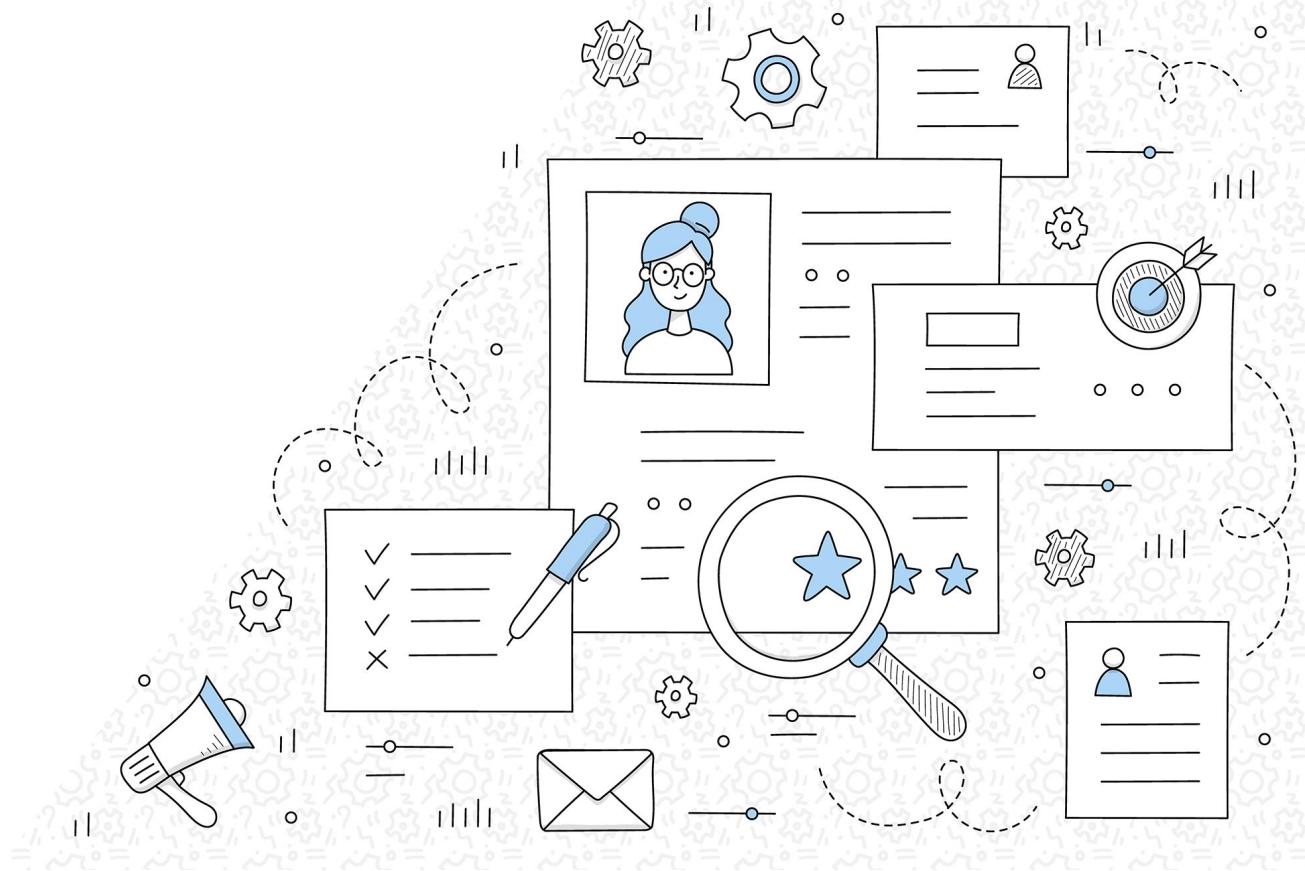
[Tesseract Google](#)

[OpenCV](#)

[PDFPLumber](#)

[Tabula](#)

[Comprendre l'OCR, le guide](#)



SOURCES