Tesseract OCR

Site du projet : https://github.com/tesseract-ocr/tesseract

Documentation : https://tesseract-ocr.github.io/

Installation sur Debian 12

```
# 20250604 Debian 12 @ tenko
sudo apt update
sudo apt install tesseract-ocr tesseract-ocr-fra tesseract-ocr-eng
tesseract --version # tesseract 5.3.0 + leptonica-1.82.0
```

Usage

Tesseract ne lit pas le format pdf en entrée, mais on peut passer par `convert`

```
convert -density 300 page.pdf page.png && tesseract page.png -l fra page.txt
```

Article extrait de : http://lesporteslogiques.net/wiki/ - **WIKI Les Portes Logiques**
Adresse : **http://lesporteslogiques.net/wiki/ressource/logiciel/tesseract/start?rev=1749042456**
Article mis à jour: **2025/06/04 15:07**