

PDF Comparator

Comparez deux fichiers PDF et voyez ce qui a changé.
Compare two PDF files and see what changed.

PDF ANCIEN / OLD

... développé en Python.

PDF NOUVEAU / NEW

... développé en Python 3.9.

Simple, rapide et en ligne de commande.
Simple, fast and command line based.

Vos fichiers restent chez vous. Aucune donnée n'est envoyée.
Your files stay on your machine. No data is sent.

100% Hors ligne 100% Offline
Traitement local Local processing
Différences claires Clear differences
Interface en CLI CLI interface
Python 3.x Python 3.x

Un outil léger pour comparer vos documents PDF en toute confidentialité. • A lightweight tool to compare your PDF documents with total privacy.

GB English | [FR Français](#)

[Python 3.x](#) [License MIT](#) [Mode Offline First](#) [YouTube @Palks_Studio](#) [LinkedIn @Palks_Studio](#)

[PALKS STUDIO](#) [WEBSITE](#)

PDF Comparator

A simple and local PDF file comparison tool built with Python 3.9.

The tool extracts the textual content of two PDF documents, lightly normalizes the data, then displays the differences detected between the two versions.

No AI, no external API, and no file upload to any third-party service.

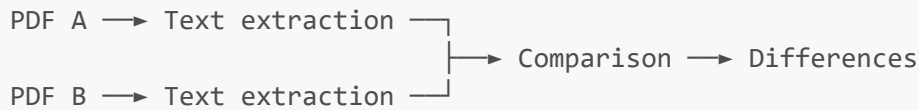
Structure

```
pdf-comparator/  
├── pdf-compare.py          → Main PDF comparison script  
├── requirements.txt        → Required Python dependencies  
├── LICENSE.md             → MIT License  
├── README.md              → English documentation  
├── README_FR.md           → French documentation  
├── docs/  
│   └── images/
```

└─ Palks_Studio.png	→ Palks Studio logo
└─ pdf_comparator.png	→ PDF Comparator presentation image

How it works

The principle is intentionally simple:



Differences are displayed directly in the terminal, showing added and removed lines.

Example:

```
--- old.pdf
+++ new.pdf

-Address: 12 Example Street
+Address: 24 Example Street

-Total: €1,200
+Total: €1,350
```

Features

- Local comparison of two PDF files
- Text extraction across all pages
- Lightweight content normalization before comparison
- Detection of added and removed lines
- No artificial intelligence
- No external API
- No document transfer
- Minimal dependency
- Command-line usage

Requirements

- Python 3
- [pypdf](#)

Installation

Clone the repository:

```
git clone REPOSITORY_URL  
cd pdf-comparator
```

Install the dependency:

```
python -m pip install -r requirements.txt
```

Usage

Run the comparator and provide the two PDF files:

```
python pdf-compare.py old.pdf new.pdf
```

The first file is considered the reference version, while the second file is the new version to compare.

If no textual difference is detected:

```
FR: Aucune différence détectée.  
EN: No differences detected.
```

Otherwise, the detected differences are displayed directly in the terminal.

Limitations

PDF Comparator compares the textual content extracted from the documents.

It does not visually compare page rendering, images, fonts, colors, or graphical layout.

The quality of the comparison also depends on the text that can actually be extracted from the PDF. A document containing only images, or a scanned document without an exploitable text layer, cannot be compared correctly.

The tool does not perform any business interpretation of the content and does not attempt to determine the meaning of the detected differences.

Privacy

Files are processed locally on the user's machine.

PDF Comparator does not use any remote service or external API, and documents are never sent to a server.

Project purpose

This project provides an intentionally lightweight solution for quickly identifying textual differences between two versions of a PDF document, without relying on an external platform or artificial intelligence system.

It can be used to review changes in documentation, reports, technical documents, or different versions of the same file.

License

This project is distributed under the MIT License.

© Palks Studio — see LICENSE.md

- <https://palks-studio.com>