

# Documento de requerimientos de software

***PDF\_READER***  
***Fecha: 16-04-2024***

## Tabla de contenido

Historial de Versiones	3
Información del Proyecto	3
Aprobaciones	3
1. Propósito	4
2. Alcance del producto	4
3. Referencias	4
4. Funcionalidades principales	4
5. Clases y características de usuarios	4
6. Entorno operativo	4
7. Requerimientos funcionales	5
7.1 extract_text_from_pdf()	5
7.2 create_dictionary()	5
7.3 count_repetitions()	5
7.4 write_file_pdf()	6
8. Reglas de negocio	6
9. Requerimientos de interfaces externas	6
9.1 Interfaces de usuario	6
9.2 Interfaces de hardware	6
9.3 Interfaces de software	6
9.4 Interfaces de comunicación	7
10. Requerimientos no funcionales	7
11. Otros requerimientos	7

### Historial de Versiones

Fecha	Versión	Autor	Organización	Descripción
31-03-2024	0.0.0	Mario Hernandez	Unitec	iniciando
09-04-2024	0.1.0	Mario Hernandez	Unitec	1er release
16-04-2024	0.2.0	Mario Hernandez	Unitec	2do release
16-04-2024	0.2.1	Mario Hernandez	Unitec	Patch 1 release 2

### Información del Proyecto

Empresa / Organización	Unitec
Proyecto	PDF-reader
Fecha de preparación	09-04-2024
Cliente	Ingeniería de sistemas
Patrocinador principal	Mario Eduardo Hernandez Montoya
Gerente / Líder de Proyecto	Mario Eduardo Hernandez Montoya
Gerente / Líder de Análisis de negocio y requerimientos	Mario Eduardo Hernandez Montoya

### Aprobaciones

Nombre y Apellido	Cargo	Departamento u Organización	Fecha	Firma

### 1. Propósito

PDF – reader v0.1.0: es proporcionar una herramienta que permita a los usuarios identificar repeticiones de frases o palabras en archivos (.pdf).

### 2. Alcance del producto

El software se extiende a cualquier usuario que necesite analizar repeticiones de texto en archivos (.pdf).

### 3. Referencias

Librerías:

- tkinter: <https://docs.python.org/3/library/tkinter.html>
- PyPDF2: <https://pypdf2.readthedocs.io/en/3.x/>
- re: <https://docs.python.org/3/library/re.html#module-re>
- FPDF: <http://www.fpdf.org/>

### 4. Funcionalidades principales:

- `extract_text_from_pdf(pdf_file: str) -> list[str]:`
- `create_dictionary (pdf_text: list[str]) -> dict[str, int]:`
- `count_repetitions(pdf_text: list[str], dict_phrases: list[str] ], dict_words: list[str]) -> dict[str, int]:`
- `write_file_pdf(dictionary: dict[str, int], output_file: str, ], sort: bool = False) -> None:`

### 5. Clases y características de usuarios

El software está diseñado para ser utilizado por cualquier usuario que necesite identificar repeticiones de frases o palabras en archivos (.pdf).

## 6. Entorno operativo

Este software esta diseñado para que se use en PC:

- Windows : 7 o posterior
- Linux: Ubuntu, Debian, entre otras.
- macOS: Snow Leopard (10.6) o posterior

Cada uno necesitara tener instalado python para poder ejecutarse

## 7. Requerimientos funcionales

### 7.1 `extract_text_from_pdf(pdf_file: str) -> list[str]:`

- **Descripción:** extraer el texto de un archivo (.pdf).
- **Prioridad:** Alta
- **Acciones iniciadoras:** la ruta de un (.pdf) en el dispositivo.
- **Comportamiento esperado:** devuelve una lista con el texto extraído de un (.pdf).
- **Requerimientos funcionales:**
  - **REQ – 1:** El texto extraído debe mantener la estructura del contenido original del (.pdf).
  - **REQ – 2:** Debe ser capaz de manejar archivos (.pdf) que contengan distintas fuentes, tamaños y estilos de texto.
  - **REQ – 3:** El texto extraído debe estar libre de errores y ser legible para el usuario.

### 7.2 `create_dictionary(pdf_text: list[str]) -> dict[str, int]:`

- **Descripción:** crea dos diccionarios uno de frases y otro de palabras de un archivo (.pdf).
- **Prioridad:** Alta

## Especificación Funcional

- **Acciones iniciadoras:** la ruta de un (.pdf) en el dispositivo.
- **Comportamiento esperado:** devuelve una lista con el texto extraído de un (.pdf).
- **Requerimientos funcionales:**
  - **REQ – 1:** Debe regresar dos diccionarios frases / palabras
  - **REQ – 2:** Debe de limpiar el texto para que las frases sean legibles

7.3 `count_repetitions(pdf_text: list[str], dict_phrases: list[str], dict_words: list[str]) -> dict[str, int]:`

- **Descripción:** cuenta la cantidad de veces que se repite una frase o palabra en un texto.
- **Prioridad:** Media
- **Acciones iniciadoras:** se necesita extraer el texto de 2 (.pdf).
- **Comportamiento esperado:** devuelve dos diccionario con las frases y palabras con sus repeticiones en el texto.
- **Requerimientos funcionales:**
  - **REQ – 1:** La función debe ser capaz de recibir como entrada una lista de texto extraído de al menos dos archivos (.pdf) para contar las repeticiones.
  - **REQ – 2:** La función debe recorrer todo el texto proporcionado para determinar las repeticiones de cada frase o palabra en el texto.

7.4 `write_file_pdf(dictionary: dict[str, int], output_file: str) -> None:`

- **Descripción:** Crea un archivo (.pdf) con la información de un diccionario.
- **Prioridad:** Media
- **Acciones iniciadoras:** Tener un diccionario y el nombre de salida del archivo.
- **Comportamiento esperado:** Crea un archivo (.pdf) con la información del diccionario.
- **Requerimientos funcionales:**
  - **REQ – 1:** La función debe escribir la información del diccionario en el archivo (.pdf) como “clave: valor” en líneas separadas.

- **REQ – 2:** La función debe recibir como entrada el nombre del archivo de salida el cual sera (.pdf).

## 8. Reglas de negocio

El programa esta hecho para leer 2 pdf y extraer su texto y comparar cuantas veces se repiten sus frases entre si por lo que le seria util a alguien que quiera saber las repeticiones de frases o palabras, posteriormente se podria desarrollar un software complemente que remplace ciertas palabras en el texto

## 9. Requerimientos de interfaces externas

**9.1 Interfaces de usuario:** Este software en particular no tiene un GUI, solo se usa la consola y el explorador de archivos de tu SO.

**9.2 Interfaces de hardware:** Soporta Computadoras.

**9.3 Interfaces de software:** python

**Librerias:**

- **tkinter:** Es usada para el subprocesso de abrir el explorador de archivos del OS.
- **PyPDF2:** Es usada para la lectura y extraccion del contenido de un (.pdf).
- **re:** Es usada para limpiar el texto y dividirlo en frases o palabras.
- **FPDF:** Es usada para la creacion de archivos (.pdf).

**9.4 Interfaces de comunicación:** Este software es de uso local y sin conexión.

## 10. Requerimientos no funcionales

- El software solamente lee archivos PDF
- El software devuelve dos archivo (.pdf).

## 11. Otros requerimientos

Es necesario tener Python instalado para ejecutar el software.

## 12. Glosario

## Especificación Funcional

- Re – “regex” básicamente **regular** **expresions**
- PDF – Formato de documento portátil.
- GUI: Interfaz Gráfica de Usuario