Documento de requerimientos de software

PDF_READER
Fecha: 16-04-2024

Tabla de contenido

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

Información del Proyecto

Empresa / Organización	Unitec
Proyecto	PDF-reader
Fecha de preparación	09-04-2024
Cliente	Ingenieria 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

Librerias:

- **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**

4. Funcionalidades principales:

- `extract_text_from_pdf(pdf_file: str) -> list[str]:`
- `count_repetitions(pdf_text: list[str], dict_phrases: list[str] ], dict_words: list[str]) -> dict[str, int]:`
- `write_file_txt(dictionary: dict[str, int], output_file: str) -> 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 extraido 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 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.

7.3 write_file_txt(dictionary: dict[str, int], output_file: str) -> None:

- **Descripción:** Crea un archivo .txt con la informacion de un diccionario.

- **Prioridad:** Media

- **Acciones iniciadoras:** Tener un dictionario y el nombre de salida del archivo.

- **Comportamiento esperado:** Crea un archivo .txt con la informacion del diccionario.

- **Requerimientos funcionales:**
- **REQ – 1:** La función debe escribir la información del diccionario en el archivo .txt como “clave: valor” en líneas separadas.
- **REQ – 2:** La función debe recibir como entrada el nombre del archivo de salida el cual sera .txt

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 palabaras, posteriormente se podria desarrollar un software complemente que remplaze 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 subproceso 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.

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 un archivo .txt

11. Otros requerimientos

Es necesario tener Python instalado para ejecutar el software.

12. Glosario

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