**Documento de requerimientos de software**

***Comparador de PDF***

***Fecha: 02/04/2024.***

***Leonardo Miranda Rico***

***Luis Ángel Ventura Rivera***

***Jorge Emilio Gutiérrez Cortés***

**Tabla de contenido**

Historial de Versiones 3

Información del Proyecto 3

Aprobaciones 3

1. Propósito 4

2. Alcance del producto / Software 4

3. Referencias 5

4. Funcionalidades del producto 5

5. Clases y características de usuarios 5

6. Entorno operativo 6

7. Requerimientos funcionales 6

7.1. Cargar Documento PDF 7

7.2. Cargar Diccionario 8

7.3. Análisis de Texto 9

7.4. Identificar y Registrar Coincidencias 9

7.5. Generar Archivo de Coincidencias 10

7.6. Interfaz de Usuario Intuitiva 10

8. Reglas de negocio 11

9. Requerimientos de interfaces externas 11

9.1. Interfaces de usuario 11

9.2. Interfaces de hardware 11

9.3. Interfaces de software 11

9.4. Interfaces de comunicación 12

10. Requerimientos no funcionales 12

11. Otros requerimientos 12

12. Glosario 13

**Historial de Versiones**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Fecha** | **Versión** | **Autor** | **Organización** | **Descripción** |
| 02/04/2024. | 1 | Leonardo Miranda Rico | UNITEC | Inicio de la documentación y desarrollo de los puntos: 1,2,4,5 y 7. |
| 09/04/2024. | 2 | Leonardo Miranda Rico | UNITEC | Desarrollo de la documentación y puntos faltantes. |

**Información del Proyecto**

|  |  |
| --- | --- |
| Empresa / Organización | UNITEC |
| Proyecto | Comparador de PDF |
| Fecha de preparación | 02/04/2024. |
| Cliente | UNITEC |
| Patrocinador principal | N/A |
| Gerente / Líder de Proyecto | N/A |
| Gerente / Líder de Análisis de negocio y requerimientos | N/A |

**Aprobaciones**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nombre y Apellido** | **Cargo** | **Departamento u Organización** | **Fecha** | **Firma** |
| N/A | N/A | N/A | N/A | N/A |

1. **Propósito**

El propósito del software es el desarrollo de una herramienta para analizar documentos de texto en formato PDF con el objetivo de identificar y extraer palabras o frases que coincidan con un diccionario dado en otro documento de texto; generando como resultado un archivo con las coincidencias encontradas.

En este documento se cubrirá la totalidad del alcance del software, la cual incluye:

* Análisis de documentos en formato PDF.
* Identificación y extracción de palabras o frases que coincidan con el diccionario.
* Generación de un archivo de salida con las palabras o frases identificadas.

1. **Alcance del producto / Software**

El software se enfocará en proporcionar las siguientes funcionalidades principales:

1. Análisis de documento PDF:

El software será capaz de recibir documentos en formato PDF para su análisis.

1. Identificación de coincidencias:

Utilizando un diccionario definido en un archivo de texto, proporcionado por el usuario, el software identificará y extraerá las palabras o frases que coincidan dentro del documento PDF.

1. Generación de Archivo de Salida:

Las palabras o frases identificadas se guardarán en un archivo de salida.

El software tiene como propósito facilitar la detección de coincidencias al usuario dentro de un archivo PDF, esto facilitara la búsqueda de información y/o detección de plagio en algún documento.

1. **Referencias**

El software fue desarrollado en el lenguaje de programación Python, con la ayuda de las librerías:

* MinerPDF: Herramienta de extracción de texto para documentos PDF.

Desarrollador: Desconocido

* Customkinter: Una biblioteca de interfaz de usuario de Python moderna y personalizable basada en Tkinter.

Desarrollador: Tom Schimansky

* CTkMessagebox: Herramienta de cuadro de mensajes moderno y totalmente personalizable para CustomTkinter.

Desarrollador: Akash Bora

1. **Funcionalidades del producto**

Las funcionalidades del software son:

1. Cargar Documento PDF
2. Cargar Diccionario
3. Análisis de Texto
4. Identificar y Registrar Coincidencias
5. Generar Archivo de Coincidencias
6. Interfaz de Usuario Intuitiva

1. **Clases y características de usuarios**

Dentro de las clases de usuarios a quienes les resultaría útil este software encontramos:

* Docentes:

Quienes requieren herramientas para analizar documentos extensos en busca de información relevante para sus investigaciones, y revisan múltiples entregas de trabajos de una gran cantidad de estudiantes y se deben cerciorarse que estos no sean plagiados.

* Estudiantes:

Quienes realizan investigaciones para proyectos académicos, y requieren herramientas para analizar información de diferentes fuentes.

1. **Entorno operativo**

El software está diseñado para ser utilizado en entornos que cumplan con los siguientes requisitos mínimos:

1. El software es compatible con los siguientes sistemas operativos:

* Windows 10 o posterior.
* macOS 10.12 (Sierra) o posterior.
* Distribuciones de Linux compatibles con las bibliotecas necesarias.

1. Se recomienda un hardware:

* De 4 GB de memoria RAM aunque se recomiendan 8 GB o más para un rendimiento óptimo.
* Espacio en disco suficiente para la instalación del software y almacenamiento de archivos PDF y txt.

Es posible que se requieran ciertas aplicaciones adicionales para el funcionamiento del software, como bibliotecas de procesamiento de texto.

El software es compatible con la carga y lectura de documentos en formato PDF, que cumplen con las especificaciones de Adobe PDF y archivos de texto con la extensión .txt.

1. **Requerimientos funcionales**

* Cargar Documento PDF:

El usuario podrá cargar un documento en formato PDF desde su sistema de archivos local para su análisis.

* Cargar Diccionario:

El usuario podrá cargar un documento con un conjunto de términos o frases clave que el software buscará dentro del documento PDF durante el análisis.

* Análisis de Texto:

Una vez que se haya cargado el documento PDF y se hayan cargado los términos del diccionario, el software analizará el contenido del documento PDF en busca de coincidencias con los términos especificados.

* Identificar y Registrar Coincidencias:

El software identificará las palabras o frases dentro del documento que coincidan con los términos del diccionario.

* Generar Archivo de Coincidencias:

Después de completar el análisis, el software generará un archivo de resultados que incluirá las palabras o frases identificadas.

* Interfaz de Usuario Intuitiva:

El software contará con una interfaz de usuario intuitiva y fácil de usar que permitirá a los usuarios cargar documentos y acceder a los resultados del análisis de manera eficiente.

**7.1 Cargar Documento PDF**

Descripción: El usuario podrá cargar un documento en formato PDF desde su sistema de archivos local para su análisis.

Prioridad: Nivel alto.

Acciones iniciadoras y comportamiento esperado: Secuencia de acciones de usuario y respuestas esperadas del sistema para esta funcionalidad.

Requerimientos funcionales: Lista detallada de los requerimientos funcionales asociados a esta funcionalidad.

Para cada requerimiento funcional se establece como debe mostrarse el software y cuales comportamientos debe desempeñar para que el usuario pueda realizar la función que necesita.

Cada requerimiento debe ser identificado unívocamente, para lo cual se recomienda usar un número de secuencia, que tenga algún significado y de formato común a toda la organización. Por ejemplo:

REQ-1:

REQ-2:

REQ-3:

**7.2 Cargar Diccionario**

Descripción: El usuario podrá cargar un documento con un conjunto de términos o frases clave que el software buscará dentro del documento PDF durante el análisis.

Prioridad: Nivel alto.

Acciones iniciadoras y comportamiento esperado: Secuencia de acciones de usuario y respuestas esperadas del sistema para esta funcionalidad.

Requerimientos funcionales: Lista detallada de los requerimientos funcionales asociados a esta funcionalidad.

Cada requerimiento debe ser identificado unívocamente, para lo cual se recomienda usar un número de secuencia, que tenga algún significado y de formato común a toda la organización. Por ejemplo:

REQ-1:

REQ-2:

REQ-3:

**7.3 Análisis de Texto**

Descripción: Una vez que se haya cargado el documento PDF y se hayan cargado los términos del diccionario, el software analizará el contenido del documento PDF en busca de coincidencias con los términos especificados.

Prioridad: Nivel alto.

Acciones iniciadoras y comportamiento esperado: Secuencia de acciones de usuario y respuestas esperadas del sistema para esta funcionalidad.

Requerimientos funcionales: Lista detallada de los requerimientos funcionales asociados a esta funcionalidad.

REQ-1:

REQ-2:

REQ-3:

**7.4 Identificar y Registrar Coincidencias**

Descripción: El software identificará las palabras o frases dentro del documento que coincidan con los términos del diccionario.

Prioridad: Nivel alto.

Acciones iniciadoras y comportamiento esperado: Secuencia de acciones de usuario y respuestas esperadas del sistema para esta funcionalidad.

Requerimientos funcionales: Lista detallada de los requerimientos funcionales asociados a esta funcionalidad.

REQ-1:

REQ-2:

REQ-3:

**7.5 Generar Archivo de Coincidencias**

Descripción: Después de completar el análisis, el software generará un archivo de resultados que incluirá las palabras o frases identificadas.

Prioridad: Nivel alto.

Acciones iniciadoras y comportamiento esperado: Secuencia de acciones de usuario y respuestas esperadas del sistema para esta funcionalidad.

Requerimientos funcionales: Lista detallada de los requerimientos funcionales asociados a esta funcionalidad.

REQ-1:

REQ-2:

REQ-3:

**7.6 Interfaz de Usuario Intuitiva**

Descripción: El software contará con una interfaz de usuario intuitiva y fácil de usar que permitirá a los usuarios cargar documentos y acceder a los resultados del análisis de manera eficiente.

Prioridad: Nivel medio.

Acciones iniciadoras y comportamiento esperado: Secuencia de acciones de usuario y respuestas esperadas del sistema para esta funcionalidad.

Requerimientos funcionales: Lista detallada de los requerimientos funcionales asociados a esta funcionalidad.

REQ-1:

REQ-2:

REQ-3:

1. **Reglas de negocio**

El software debe ser un analizador de textos, que al pasarle un archivo de texto en formato PDF, identifique palabras o frases tomadas de otro archivo, las que coincidan las guarde en un archivo de salida.

* El encargado de la configuración de repositorio de código es Luis Ángel Ventura Rivera.
* El encargado de la creación de código es Jorge Emilio Gutiérrez Cortes
* El encargado de la creación de documentación es Leonardo Miranda Rico.

El proyecto será desarrollado en equipos de 3 personas.

El proyecto se presentará 2 clases antes del examen final.

1. **Requerimientos de interfaces externas**
2. **Interfaces de usuario**

El software debe proporcionar una interfaz de usuario intuitiva y fácil de usar que permita a los usuarios cargar documentos PDF y archivos de texto así como ver los resultados de análisis.

1. **Interfaces de hardware**

El software debe ser compatible con dispositivos de entrada estándar, como teclados y ratones, para permitir a los usuarios interactuar con la interfaz de usuario y realizar acciones dentro del software.

Debe ser compatible con monitores de pantalla estándar para garantizar una visualización adecuada de la interfaz de usuario y los resultados del análisis.

1. **Interfaces de software**

El software debe ser capaz de interactuar con el sistema de archivos local del usuario para permitir la carga y guardado de documentos en formato PDF y .txt.

Debe ser compatible con múltiples plataformas, incluyendo Windows, macOS y distribuciones de Linux, para garantizar la accesibilidad.

1. **Interfaces de comunicación**

Se debe proporcionar documentación clara y soporte técnico para ayudar a los usuarios a comprender y utilizar eficazmente el software, a través de recursos en línea, tutoriales y soporte.

1. **Requerimientos no funcionales**

* Rendimiento:

El software debe ser capaz de analizar documentos PDF de tamaño moderado en un tiempo razonable, menos de 1 minuto por documento.

* Usabilidad:

La interfaz de usuario debe ser intuitiva y fácil de usar, permitiendo a los usuarios realizar tareas sin dificultad.

* Seguridad:

El software debe proteger la privacidad y la integridad de los datos del usuario, garantizando que la información confidencial no sea comprometida durante el análisis de documentos.

1. **Otros requerimientos**

* Desempeño:

Especifica los niveles de desempeño esperados del software.

* Legales y Regulatorios:

Identifica y asegura que el software cumpla con todas las leyes y regulaciones aplicables, como la protección de datos personales, los derechos de autor y cualquier otra normativa específica de la industria o región.

* Pruebas:

Define los criterios y procedimientos de prueba que se utilizarán para validar la funcionalidad y el desempeño del software, incluyendo pruebas unitarias, de integración, de sistema y de aceptación del usuario.

1. **Glosario**

* Análisis de texto: Proceso de examinar y comprender el contenido de un documento de texto con el objetivo de extraer información relevante.
* Documento PDF: Formato de archivo portátil que conserva el diseño original del documento, incluidos textos, gráficos y elementos de formato, independientemente del software, hardware y sistema operativo utilizado para verlo o imprimirlo.
* Archivo de texto: Es un tipo de archivo de computadora que contiene texto sin formato, es decir, caracteres legibles por humanos.
* Requerimientos funcionales: Especificaciones detalladas de las funciones y acciones que un sistema debe ser capaz de realizar.
* Requerimientos no funcionales: Requisitos que describen atributos de calidad o restricciones que afectan al sistema, pero no están relacionados directamente con las funcionalidades específicas que el software debe realizar.
* Python: Python es un lenguaje de programación que te permite trabajar rápidamente e integrar sistemas de manera más efectiva.
* Librería: Colección de código desarrollado previamente que los programadores pueden utilizar para desarrollar software de manera más ágil.