

Documento de requerimientos de software

PDFTextAnalyzer

Fecha: 02/03/2024

Tabla de contenido

Historial de Versiones	2
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	
6. Entorno operativo	7
7. Requerimientos funcionales	8
9.1. (Nombre de la funcionalidad 1)	
9.2. (Nombre de la funcionalidad 2)	8
9.3. (Nombre de la funcionalidad N)	
8. Reglas de negocio	9
9. Requerimientos de interfaces externas	9
9.1. Interfaces de usuario	9
9.2. Interfaces de hardware	
9.3. Interfaces de software	9
9.4. Interfaces de comunicación	10
10. Requerimientos no funcionales	10
11. Otros requerimientos	10
12. Glosario	11

Historial de Versiones

Fecha	Versión	Autor	Organización	Descripción
2 de Marzo	Patch-1	Rodrigo Franco Piña	Unitec Atizapán	Se implemento la primera parte del código (incompleto)
9 de Abril	Patch-2	Rodrigo Franco Piña	Unitec Atizapán	Termino la primera fase del código
12 de Abril	Patch-3	Rodrigo Franco Piña	Unitec Atizapán	Se añadió la opción de poder insertar cualquier PDF
14 de Abril 10:47 a.m	Patch-4	Rodrigo Franco Piña	Unitec Atizapán	Se corrigió el error indeterminado y espacios en blanco que en ocasiones generaba el PDF
14 de Abril 21:40	Patch5-Final	Rodrigo Franco Piña	Unitec Atizapán	Se subió como archivo el código

Información del Proyecto

Empresa / Organización	Unitec Atizapán
Proyecto	Repositorio en GitHub (comparador de PDFs)
Fecha de preparación	12 de Marzo
Cliente	Gerardo de la Rosa
Patrocinador principal	Unitec Atizapán
Gerente / Líder de Proyecto	Alma Vianney Damian Galan
Gerente / Líder de Análisis de negocio y requerimientos	Alma Vianney Damian Galan

Aprobaciones

Nombre y Apellido	Cargo	Departamento u Organización	Fecha	Firma
Itzel Beatriz Juárez Marin	Documentación	Departamento de Modelo de Diseño en software	14/04/2024	
Alma Vianney Damian Galan	Desarrollador de software	Departamento de Desarrollo de Software	14/04/2024	
Rodrigo Franco Piña	Administrador de Repositorio en GitHub	Departamento de Administrador de Repositorio	14/04/2024	

1. Propósito

El documento define que el producto consistirá en un analizador de textos especialmente creado para manejar archivos en formato PDF. Este analizador será capaz de reconocer palabras o frases específicas provenientes de un archivo de referencia y luego guardarlas en un archivo de salida designado el cual le permitirá al usuario final realizar un mejor análisis de la información consultada.

El analizador no realizará ningún tipo de traducción o modificación del contenido del archivo PDF original, solo identificará y almacenará las palabras o frases clave especificadas, con esto el usuario también decidirá como procesar la información almacenada y ejecutará un análisis según su objetivo.

2. Alcance del producto / Software

El propósito es automatizar la identificación y extracción de información relevante de archivos PDF, facilitando la consulta en base a criterios de búsqueda específicos para su fácil análisis de datos para los usuarios.

Los beneficios que brinda este proyecto es la eficiencia en la Gestión de Información, mejora la productividad, proporciona acceso rápido a información contenida en documentos PDF, facilita la recuperación de información específica de manera rápida y eficiente. Este tipo de herramienta sería útil en diversas situaciones, incluyendo:

Investigación académica: Los estudiantes y académicos podrían utilizar este software para analizar documentos relevantes y extraer las palabras clave más importantes, lo que les ayudaría a comprender mejor el contenido y a identificar temas relevantes para sus investigaciones.

Análisis de mercado: Las empresas podrían utilizar este tipo de herramienta para analizar informes de mercado y extraer palabras clave relacionadas con tendencias, competidores, o productos específicos. Esto les ayudaría a tomar decisiones informadas sobre estrategias de marketing y desarrollo de productos.

Análisis de competencia: Las empresas también podrían utilizar el software para analizar documentos de la competencia y extraer palabras clave relacionadas con sus productos, estrategias de marketing, y otros aspectos relevantes. Esto les proporcionaría información valiosa para mejorar su propia estrategia competitiva.

Análisis de opiniones: Los investigadores de opinión pública o empresas de análisis de datos podrían utilizar este software para analizar documentos que contienen opiniones de clientes o usuarios. La extracción de palabras clave y frases relevantes les permitiría identificar tendencias y patrones en las opiniones de los usuarios.

Por lo que sería un software que analiza palabras clave y frases en documentos PDF puede ser una herramienta valiosa para diversas aplicaciones, ayudando a los usuarios a extraer información importante y a realizar análisis más eficientes y precisos.

Objetivos y metas: **Automatizar la Extracción de Información Relevante:** Desarrollar algoritmos de procesamiento de texto que identifiquen y extraigan información específica sean palabras o frases clave de manera automática. Para proporcionar acceso rápido y preciso a datos relevantes, lo que puede mejorar el análisis de la información consultada con una gran eficacia y autonomía.

3. Referencias

Referencias:

"Software Engineering: A Practitioner's Approach" de Roger S. Pressman

Published by McGraw-Hill, a business unit of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020. Copyright © 2010 by The McGraw-Hill Companies, Inc.

4. Funcionalidades del producto

Procesamiento de Archivos PDF:

El analizador debe ser capaz de cargar archivos en formato PDF como entrada.

Identificación de Palabras o Frases Clave:

Funcionalidad para analizar el contenido del archivo PDF y buscar palabras o frases específicas proporcionadas en el diccionario

La capacidad de tomar las palabras o frases clave de otro archivo como referencia.

Comparación y Coincidencia:

Algoritmos para comparar las palabras o frases identificadas en el archivo PDF con las palabras o frases de referencia en el diccionario y determinar si hay coincidencias.

Almacenamiento de Resultados:

Funcionalidad para guardar las palabras o frases coincidentes en un archivo de salida.

Estas funcionalidades permitirían al analizador de textos cumplir con el objetivo que es identificar y extraer palabras o frases específicas de archivos PDF y guardarlas en un archivo de salida, facilitando así la gestión y análisis de información.

5. Clases y características de usuarios

Clasificación por Frecuencia de Uso:

Usuarios Regulares: Utilizan el producto de manera frecuente como parte de sus tareas habituales.

Usuarios Ocasionales: Utilizan el producto de manera ocasional, no tan frecuentemente como los usuarios regulares.

Clasificación por Grupo de Funcionalidades Utilizadas:

Usuarios Básicos: Utilizan las funcionalidades principales del producto para llevar a cabo tareas básicas.

Usuarios Avanzados: Utilizan funcionalidades adicionales o avanzadas del producto para realizar tareas más complejas o específicas.

Clasificación por Privilegios de Seguridad:

Usuarios Normales: Tienen acceso estándar al sistema y pueden realizar acciones permitidas dentro de su rol.

Usuarios - Administradores: Tienen privilegios elevados y pueden acceder a funciones de administración, configuración y gestión de usuarios.

Clasificación por Nivel de Experiencia:

Usuarios Principiantes: Tienen poca o ninguna experiencia previa con el producto, pueden necesitar orientación adicional.

Usuarios Experimentados: Tienen experiencia previa con el producto o tecnologías, aplicaciones similares y pueden utilizar el producto de manera más autónoma.

6. Entorno operativo

Computadoras de escritorio o portátiles: Los usuarios pueden ejecutar el analizador de textos en computadoras estándar.

Otros Sistemas o Componentes:

Software de Visualización de PDF: Es posible que el analizador de textos dependa de software adicional para la visualización y manipulación de archivos PDF.

7. Requerimientos funcionales

- 1- Carga de diccionario de palabras
- 2- Alta de archivos en formato PDF
- 3- Validación de tamaño del archivo
- 4- Validación de formato del archivo
- 5- Especificación de criterios de búsqueda
- 6- Proceso de escaneo en archivos y validación de la información
- 7- Generación de reporte con las palabras encontradas
- 8- Consulta de archivo generado
- 9- Descarga de archivo

Funcionalidad 1 (Selección de PDFs):

Descripción: Se selecciona el PFD al leer y el diccionario para poder mostrar las palabras y frases en común

Funcionalidad 2 (Resultado):

El software generara las palabras y frases en común que tienen el diccionario y el PDF a leer, generando así un archivo en .txt como resultado

REQ-1: PDF con texto y dirección en el ordenador

REQ-2: PDF con palabras y/o frases enlistadas

REQ-3: Python instalado

8. Reglas de negocio

Gestión de Archivos:

Los usuarios solo podrán cargar archivos en formato PDF al analizador de textos.

Los archivos cargados deben cumplir con ciertos requisitos de tamaño y formato.

Identificación de Palabras o Frases Clave:

Solo se identificarán como palabras o frases clave aquellas que coincidan exactamente con los criterios de búsqueda especificados por el diccionario.

Almacenamiento de Resultados:

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

Los resultados almacenados deben ser precisos y reflejar fielmente las palabras o frases clave identificadas.

9. Requerimientos de interfaces externas

9.1. Interfaces de usuario

1. El usuario insertara los PDFs escribiendo la dirección en donde se encuentra
2. El usuario generara un archivo .txt como resultado y deberá de encontrarlo en el apartado de acceso rápido

9.2. Interfaces de hardware

El software podrá ser ejecutado en cualquier dispositivo que cuente con un ordenador virtual con acceso a Python

9.3. Interfaces de software

Nuestro software interactúa principalmente con la red de archivos locales en el ordenador ejecutado

9.4. Interfaces de comunicación

Como protocolo de mensajería decidimos mandar y recibir las actualizaciones de nuestros códigos mediante WhatsApp de igual manera, de igual forma no requerimos de seguridad en las comunicación y encriptación.

Y para finalizar documentamos nuestro avance en Google Docs.

10. Requerimientos no funcionales

- 1- Usabilidad
- 2- Legibilidad de código
- 3- Documentación clara y concisa
- 4- Modularidad
- 5- Abstracción adecuada
- 6- Manejo de Errores
- 7- Eficiencia y rendimiento
- 8- Seguridad

11. Otros requerimientos

Requerimientos de red: Si el software necesita acceder a recursos o servicios en red, se deben especificar los requerimientos de conectividad, ancho de banda, seguridad de la red, etc.

Requerimientos de seguridad: Políticas y medidas de seguridad que deben implementarse para proteger los datos procesados por el software, incluyendo el cifrado de datos, la autenticación de usuarios, el control de acceso, etc.

Requerimientos de escalabilidad: Si se espera que el software maneje un aumento en la carga de trabajo o el número de usuarios en el futuro, se deben considerar

los requerimientos de escalabilidad para garantizar que el sistema pueda crecer y adaptarse según sea necesario.

Requerimientos de capacitación: Capacitación necesaria para que los usuarios finales puedan utilizar el software de manera efectiva y eficiente.

12. Glosario

Analizador de textos: Sistema de software diseñado para examinar y procesar archivos de texto, identificando palabras o frases específicas y realizando acciones basadas en ciertos criterios.

Archivo de texto en formato PDF: Documento digital que sigue el formato de archivo (PDF), utilizado para presentar documentos de texto de manera independiente del software, hardware y sistema operativo.

Plataforma de análisis: Entorno informático donde se ejecuta el software de análisis de texto y se procesan los archivos PDF de entrada.

Archivo de referencia: Documento que contiene las palabras o frases que se buscarán en el texto PDF de entrada para identificar coincidencias.

Archivo de salida: Documento generado por el analizador de textos que contiene las palabras o frases identificadas como coincidencias entre el texto PDF de entrada y el archivo de referencia.

Coincidencia: Ocurre cuando una palabra o frase presente en el archivo de texto PDF de entrada coincide con las palabras o frases definidas en el archivo de referencia.

Algoritmo de coincidencia: Método o conjunto de reglas utilizadas por el analizador de textos para determinar si una palabra o frase en el archivo de texto PDF de entrada coincide con las definidas en el archivo de referencia.

Extracción de texto: Proceso mediante el cual el contenido textual se extrae del archivo PDF para su posterior análisis por parte del software.

Desarrollo: Proceso de creación e implementación del software, que implica la codificación, pruebas y ajustes necesarios para asegurar su funcionamiento adecuado.

Diseño: Etapa del proceso de desarrollo de software que involucra la definición de la arquitectura y estructura del sistema, así como la planificación de su funcionamiento y comportamiento.

Puesta en producción: Fase en la que el software es implementado y puesto en funcionamiento en un entorno real

Requerimientos de software: Documento que describe las funcionalidades, características y restricciones que debe cumplir el software a desarrollar, basado en las necesidades del cliente o usuario final.

Este glosario proporciona una lista de términos y sus significados específicos que se aplican directamente al proyecto de analizador de textos. Esto es importante porque establece un conjunto común de definiciones que todos los miembros del equipo de desarrollo y las partes interesadas en el proyecto pueden utilizar y entender de la misma manera.

Con la finalidad de evitar malentendidos y confusiones sobre el lenguaje técnico utilizado durante el desarrollo del proyecto. Esto garantiza que todos estén en la misma página y compartan una comprensión consistente de los conceptos clave, los requisitos del proyecto y los objetivos a alcanzar.