



MANUAL TÉCNICO

SENASoft – Vertical 2
Jeison Stanlin Zabala Castrillón
Jaiber Santiago Montoya Garzón
Bryan Alexis Guzmán Bermúdez



22 DE OCTUBRE DE 2021
SENA – CENTRO DE CONFECCIÓN Y MODA
Calle 63 No. 58 B - 03 Itagüí

Tabla de contenidos:

- [Prerrequisitos.](#)
- [Tecnologías utilizadas:](#)
 - [Hardware de los equipos.](#)
 - [Lenguajes de programación y frameworks.](#)
 - [Herramientas de desarrollo.](#)
 - [Herramientas externas.](#)
- [Docker.](#)
- [Uso.](#)
- [Información de utilidad.](#)

Prerrequisitos.

Son los requisitos previos para la correcta ejecución del nuestro software.

- **Docker:** Es el software que ejecutará el contenedor creado para nuestro software, en cuanto a su versión, es ideal usar la más reciente.
- **Navegador web:** Docker nos creará contenedores con puertos en específico, a los cuales accederemos a través del navegador.

Tecnologías utilizadas.

En este apartado hablaremos sobre los programas, tecnologías y herramientas utilizadas para el desarrollo de nuestro software.

Hardware de los equipos:

- **Sistema operativo:** Windows 10 Enterprise.
- **Procesador:** AMD Ryzen 7 3700U with Radeon Vega Mobile Gfx.
- **Memoria RAM:** 16 GB.
- **Disco duro:** Intel SSD 512 GB.

Lenguajes de programación y frameworks:

- Python 3.
- Java JDK 14.
- Spring Boot con Gradle.

Herramientas de desarrollo:

- Visual Studio Code.
- Eclipse con extensión de Spring Boot.
- Docker Desktop.

Herramientas externas:

- **Tesseract OCR:** Es un motor de reconocimiento óptico de caracteres, es software libre, como su descripción indica, es usado para el reconocimiento de texto en imágenes, es capaz de reconocer muchos idiomas, pero se le debe indicar en cuál se va a trabajar. Lo usamos para la clasificación de los PDFs formados por imágenes.
- **Poppler:** Es una biblioteca de software libre, la cual tiene la capacidad de gestionar el contenido de los PDFs, en nuestro caso, lo usamos como una herramienta para obtener la primera página de un PDF.

Docker.

Enlace a los repositorios de Docker-Hub: <https://hub.docker.com/u/jsmontoya05>

Uso.

En este apartado hablaremos sobre el uso correcto de nuestro software, la idea es aclarar los pasos que se deben de seguir.

1. Ingresar a “http://localhost:8080/APISERVICE/verOrigen”.
2. Los archivos PDFs que se desean clasificar, deben ir en la carpeta “pdfs” que se ubica dentro de una carpeta, que en este caso se llama “app”.
3. Dirígete a tu navegador de preferencia e introduce el siguiente enlace:
“http://localhost:8080/APISERVICE/moveFile”, deberá informar que el proceso se realizó.
4. Se puede verificar el funcionamiento accediendo a la consola o mirando un servicio de prueba llamado: “http://localhost:8080/APISERVICE/verSinClasificar”.

Información de utilidad.

En este apartado hablaremos de información que podría resultar útil a la hora de usar nuestro software.

- El software clasifica entre PDFs cuyo contenido es texto y los PDFs cuyo contenido es imagen.
- El tiempo de respuesta con los PDFs cuyo contenido es texto es mucho menor a los PDFs cuyo contenido son imágenes, de igual forma, estos últimos no superan los 5 segundos por PDF.
- A la hora de evaluar un PDF y no ser capaz de rescatar texto, pasará a tratarlo como imagen, lo cual generará una imagen temporal para escanear el texto que pueda tener. En caso de no rescatar nada de esta forma, se enviará a la categoría de “Vacíos”.
- Al finalizar la ejecución de nuestro software, se creará un archivo log.txt con el año y mes actuales (Ej. “log-2021-10.txt”), en caso de ya existir, lo usará. Se entiende entonces, que habrá un archivo log.txt por cada mes, el cuál guardará todas las ejecuciones realizadas en ese mes en específico (Cada ejecución lleva fecha y hora exacta, en caso de las clasificaciones de los PDFs, los tiempos de respuesta).