

Introducción

En el proceso de desarrollo de nuestro proyecto, hemos estructurado un Product Backlog que define y organiza las tareas esenciales para la creación de un sistema de extracción automática de información de documentos PDF.

Este Product Backlog refleja la planificación inicial del proyecto, delineando las actividades clave que contribuirán a la consecución de nuestros objetivos.

El Product Backlog no representa un mapa detallado si no una guía de nuestro viaje, que deberá irse desarrollando en mayor detalle con el paso de las iteraciones, y el descubrimiento de las necesidades del sistema.

Product backlog

Cod.	Nombre y descripción	Horas	desarrolla?
001	Análisis y definición inicial de requisitos En esta actividad se realiza un análisis y una definición inicial del sistema completo, dejando paso al análisis detallado de cada componente para más adelante.	10.0	SI
002	Diseño de la arquitectura del sistema de generación de texto. En esta actividad se diseñarán todas las clases e interfaces que intervienen en este componente, así como sus entradas y salidas. Depende de 001	6.0	SI
003	Diseño de la arquitectura del sistema de lectura de texto. En esta actividad se diseñarán todas las clases e interfaces que intervienen en este componente, así como sus entradas y salidas. Depende de 001	8.0	SI

004	<p>Diseño e implementación del sistema de pruebas automáticas.</p> <p>En esta actividad se realizará la configuración de los procesos de pruebas automáticas.</p> <p>Depende de 001</p>	4.0	SI
005	<p>Diseño e implementación del sistema basado en contenedores.</p> <p>En esta actividad se desarrolla una estructura de contenedores y la comunicación entre ellos de forma que pueda ejecutarse el proyecto sin la necesidad de instalar todas las utilidades requeridas.</p> <p>Depende de 001</p>	10.0	NO
006	<p>Diseño e implementación del sistema de registro de logs.</p> <p>En esta actividad se desarrolla un sistema de registro de logs, que permita la depuración de errores, así como la obtención de información estadística de uso.</p> <p>Depende de 001</p>	4.0	NO
007	<p>Definición de un conjunto de datos de prueba.</p> <p>En esta actividad se desarrollará un conjunto de datos de pruebas con algún tipo de documento, generando manualmente la salida esperada.</p> <p>Depende de 001</p>	20.0	SI
008	<p>Implementación de un proceso automático de verificación</p> <p>En esta actividad se ejecutará el conjunto de datos de prueba sobre el sistema y se obtendrá como resultado un grado de precisión sobre el resultado esperado.</p> <p>Depende de 007</p>	8.0	SI
009	<p>Implementación de un procesador de texto basado en pdf2text.</p> <p>En esta actividad se desarrollará un procesador</p>	12.0	SI

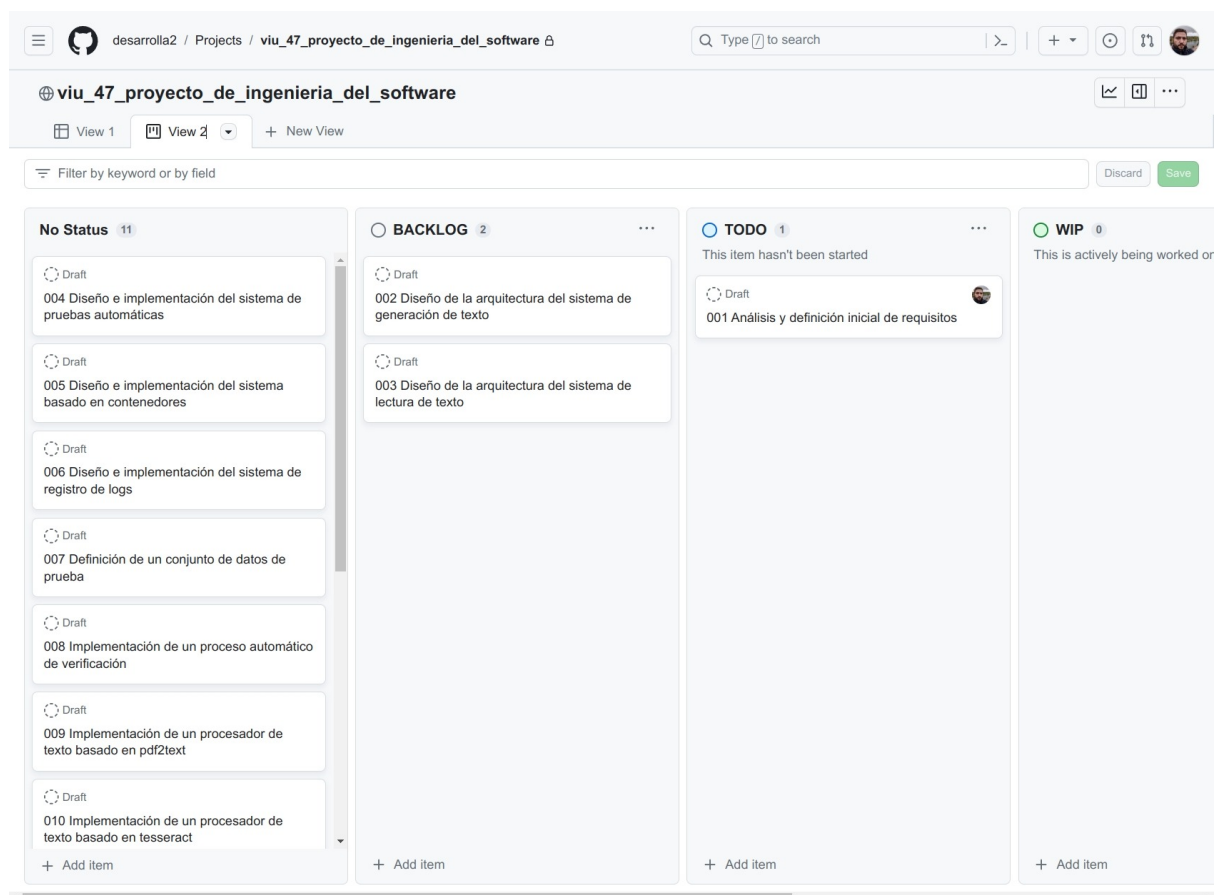
	de texto basado en pdf2text. Depende de 002		
010	Implementación de un procesador de texto basado en tesseract. En esta actividad se desarrollará un procesador de texto basado en tesseract. Depende de 002	16.0	NO
011	Investigación sobre el estado del arte de las herramientas de lectura automática de texto. En esta actividad se realizará una investigación sobre las herramientas existentes en este campo, priorizando herramientas open source sobre herramientas con licencia privativa. Depende de 001.	40.0	SI
012	Implementación de un módulo de lectura automática de texto. Como resultado de la actividad anterior se realizará la implementación con al menos una de las herramientas investigadas. Depende de 010.	40.0	SI
013	Elaboración de la documentación requerida para la asignatura.	10.0	SI
014	Publicación del proyecto con una licencia Open Source Como resultado de esta actividad se realizará los pasos necesarios para ser publicados como un proyecto open source, selección de la licencia, elaboración de la documentación, manuales de instalación, ejecución, ejecución de las pruebas, guía de estilo, guía de cómo contribuir, etc.	16.0	NO
015	Desarrollo de una interfaz de línea de comandos Como resultado de esta actividad se podrá interactuar con el sistema mediante llamadas en la línea de comandos	16.0	SI
016	Desarrollo de una interfaz web sencilla	24.0	SI

	Como resultado de esta actividad se podrá interactuar con el sistema mediante una página web, en la que se podrán arrastrar documentos y ver el resultado del análisis.		
	Total	244	

Planificación de actividades

A pesar de que se ha definido una metodología ágil donde prima la obtención de software funcionando, sobre el seguimiento de un plan minuciosamente definido, se han definido las siguientes actividades principales.

El seguimiento del proyecto se realizará a través de la herramienta de github projects en la siguiente URL <https://github.com/users/desarrolla2/projects/1/views/2>.



Sprints y entregables

Si bien en una metodología ágil debe primar la realización de un trabajo que aporte el máximo valor en cada momento, se ha diseñado el siguiente plan de trabajo que deberá ser reevaluado cada semana y adaptado a los avances realizados.

Al principio de cada sprint se realizará un análisis del trabajo a realizar, profundizando en la definición de cada una de las tareas.

Al final de cada sprint, se hará una reevaluación del calendario de trabajo en función de si se están cumpliendo los objetivos.

Sprint #0, Inicio 26/10/23, Final 09/11/23

Se han definido las siguientes actividades a realizar.

Cod.	Nombre y descripción	Horas
001	Análisis y definición inicial de requisitos	10.0
002	Diseño de la arquitectura del sistema de generación de texto	6.0
	Total	16.0

Sprint #1, Inicio 10/11/23, Final 23/11/23

Se han definido las siguientes actividades a realizar.

Cod.	Nombre y descripción	Horas
003	Diseño de la arquitectura del sistema de lectura de texto.	8.0
004	Diseño e implementación del sistema de pruebas automáticas.	4.0
007	Definición de un conjunto de datos de prueba.	10.0
	Total	22.0

Sprint #2, Inicio 24/11/23, Final 07/12/23

Se han definido las siguientes actividades a realizar.

Cod.	Nombre y descripción	Horas
------	----------------------	-------

008	Implementación de un proceso automático de verificación.	10.0
009	Implementación de un procesador de texto basado en pdf2text.	12.0
Total		22.0

Sprint #3, Inicio 08/12/23, Final 22/12/23

Se han definido las siguientes actividades a realizar.

Cod.	Nombre y descripción	Horas
011	Investigación sobre el estado del arte de las herramientas de lectura automática de texto	40.0
Total		40.0

Sprint #4, Inicio 22/12/23, Final 11/01/24

Se han definido las siguientes actividades a realizar.

Cod.	Nombre y descripción	Horas
012	Implementación de un módulo de lectura automática de texto	40.0
Total		40.0

Sprint #5, Inicio 12/01/24, Final 25/01/24

Se han definido las siguientes actividades a realizar.

Cod.	Nombre y descripción	Horas
015	Desarrollo de una interfaz de línea de comandos	16.0
Total		16.0

Sprint #6, Inicio 26/01/24, Final de la asignatura

Se han definido las siguientes actividades a realizar.

Cod.	Nombre y descripción	Horas
013	Elaboración de la documentación requerida para la asignatura	10.0
016	Desarrollo de una interfaz web sencilla	24.0

	Total	34.0
--	--------------	-------------

Dedicación

Con el fin de garantizar el progreso adecuado de este proyecto, se han planificado entregas cada dos semanas.

Se estima que el esfuerzo requerido oscila entre 16 y 40 horas por entrega, equivalente a un rango semanal de 8 a 20 horas.

Aunque esta carga de trabajo es significativa, vale la pena señalar que puedo aprovechar todo este trabajo para mi proyecto de fin de grado, lo que lo convierte en una oportunidad valiosa para mi desarrollo académico y profesional.