



Instituto  
Europeo  
de Posgrado

# Sistema de Consulta Inteligente sobre Documentos mediante RAG

|                                        |                                                   |
|----------------------------------------|---------------------------------------------------|
| Trabajo fin de estudio presentado por: | Victor Guerra Rubio y Jose Alberto Gallardo Pérez |
| Tipo de trabajo:                       | TFM IEP                                           |
| Director/a:                            |                                                   |
| Fecha:                                 | Marzo 2026                                        |

## Índice

|                                                                                      |           |
|--------------------------------------------------------------------------------------|-----------|
| <b>1. Resumen</b>                                                                    | <b>4</b>  |
| <b>2. Abstract</b>                                                                   | <b>4</b>  |
| <b>3. Introducción</b>                                                               | <b>5</b>  |
| 3.1 Objetivo General                                                                 | 6         |
| 3.2 Objetivos Específicos                                                            | 6         |
| 3.3 Metodología                                                                      | 6         |
| 3.3.1 Identificación y Análisis del Problema                                         | 6         |
| 3.3.2 Planificación de la Solución                                                   | 7         |
| 1. Ingesta de documentos                                                             | 7         |
| 2. Segmentación del contenido (Chunking)                                             | 7         |
| 3. Indexación semántica                                                              | 7         |
| 4. Recuperación de información                                                       | 7         |
| 5. Reranking de resultados                                                           | 7         |
| 6. Generación de respuestas                                                          | 7         |
| 7. Modo examen                                                                       | 8         |
| <b>4. Identificación y descripción de un problema al interior de la organización</b> | <b>8</b>  |
| 4.1 Descripción de la empresa a intervenir                                           | 8         |
| 4.1.1 Misión                                                                         | 9         |
| 4.1.2 Visión                                                                         | 9         |
| 4.1.3 Mapa de procesos                                                               | 9         |
| 4.1.4 Descripción del proceso a intervenir                                           | 10        |
| 4.1.5 Descripción del problema que presenta el proceso                               | 11        |
| <b>5. Establecimiento de planes de acción para la solución del problema</b>          | <b>11</b> |
| 5.1 Establecimiento de objetivos                                                     | 12        |
| Objetivo general                                                                     | 12        |
| Objetivos específicos                                                                | 12        |
| 5.2 Descripción de la solución                                                       | 12        |
| Interfaz de usuario                                                                  | 12        |
| Procesamiento documental                                                             | 13        |
| Motor de recuperación de información (RAG)                                           | 13        |

|                                                                                               |           |
|-----------------------------------------------------------------------------------------------|-----------|
| Funcionalidades avanzadas                                                                     | 13        |
| 5.3 Secuencia de actividades para la ejecución de la solución                                 | 13        |
| Fase 1: Base de la aplicación e ingestión documental                                          | 13        |
| Fase 2: Indexado semántico y recuperación de información (RAG básico)                         | 13        |
| Fase 3: Mejora de calidad: reranking, citas y evaluación                                      | 14        |
| Fase 4: Implementación del modo examen y cierre del proyecto                                  | 14        |
| 5.4 Descripción de recursos necesarios para la ejecución de la solución                       | 15        |
| Recursos tecnológicos                                                                         | 15        |
| Recursos humanos                                                                              | 15        |
| Recursos temporales                                                                           | 15        |
| Recursos de información                                                                       | 15        |
| 5.5 Tabla por fases                                                                           | 16        |
| <b>6. Definición del alcance de la solución, partes interesadas, cronograma y presupuesto</b> | <b>17</b> |
| 6.1 Alcance de la solución                                                                    | 17        |
| Objetivo del alcance                                                                          | 17        |
| Resultados esperados                                                                          | 17        |
| Límites del alcance                                                                           | 17        |
| 6.2 Identificación de responsables de las actividades para ejecución                          | 17        |
| Tabla 1. Roles y responsabilidades                                                            | 17        |
| 6.3 Cronograma de ejecución                                                                   | 18        |
| Hitos clave                                                                                   | 18        |
| Tabla 2. Cronograma de actividades                                                            | 18        |
| 6.4 Presupuesto                                                                               | 19        |
| Justificación del presupuesto                                                                 | 19        |
| Tabla 3. Presupuesto                                                                          | 19        |
| 6.5 Evaluación y seguimiento                                                                  | 19        |
| <b>7. Referencias bibliográficas</b>                                                          | <b>20</b> |

## 1. Resumen

Este Trabajo de Fin de Máster aborda el problema de la dificultad para localizar información específica dentro de documentos extensos en formato digital, una situación frecuente en numerosas organizaciones donde el conocimiento se encuentra disperso en informes, manuales y documentación técnica.

Para dar respuesta a esta problemática, se propone el desarrollo de un sistema inteligente basado en la arquitectura Retrieval-Augmented Generation (RAG), que permitirá a los usuarios interactuar con documentos mediante lenguaje natural. El sistema facilitará la carga de archivos y la realización de consultas, generando respuestas precisas apoyadas en fragmentos relevantes del documento original e indicando referencias como páginas o secciones.

Asimismo, el proyecto analizará distintas estrategias de optimización de la recuperación de información, incluyendo métodos de segmentación de documentos, variaciones en el número de resultados recuperados (top-k) y técnicas de reranking para mejorar la relevancia de las respuestas.

Finalmente, se incorporará una funcionalidad adicional orientada al aprendizaje denominada "modo examen", que permitirá generar automáticamente preguntas a partir del contenido del documento para evaluar la comprensión del usuario.

### Palabras clave:

Recuperación de Información, Generación Aumentada por Recuperación (RAG), Procesamiento del Lenguaje Natural, Modelos de Lenguaje, Embeddings, Bases de Datos Vectoriales, Gestión del Conocimiento, Búsqueda Semántica, Inteligencia Artificial Aplicada

## 2. Abstract

This Master's Thesis addresses the challenge of locating specific information within large digital documents, a common issue in many organizations where knowledge is stored in reports, manuals, and technical documentation.

To solve this problem, the project proposes the development of an intelligent system based on the Retrieval-Augmented Generation (RAG) architecture, enabling users to interact with documents using natural language. The system will allow users to upload files and ask questions about their content, generating accurate responses supported by relevant document fragments and including references such as page numbers or sections.

Additionally, the project will evaluate different strategies to improve information retrieval performance, including document segmentation methods, variations in the number of retrieved results (top-k), and reranking techniques to enhance response relevance.

Finally, the application will incorporate an additional learning-oriented feature called "exam mode," which will automatically generate questions based on the document content to help users assess their understanding.

#### Keywords:

Information Retrieval, Retrieval-Augmented Generation (RAG), Natural Language Processing, Language Models, Embeddings, Vector Databases, Knowledge Management, Semantic Search, Applied Artificial Intelligence

### 3. Introducción

En el entorno organizacional actual, gran parte del conocimiento corporativo se almacena en documentos digitales como informes, manuales operativos y documentación técnica, habitualmente en formato PDF. Aunque esta digitalización facilita la conservación de la información, la localización de datos concretos dentro de documentos extensos continúa siendo un proceso lento e ineficiente.

Los métodos tradicionales de búsqueda obligan a los usuarios a revisar manualmente grandes volúmenes de contenido o a realizar búsquedas por palabras clave que no siempre ofrecen resultados precisos. Esta situación genera pérdida de tiempo, reduce la productividad y dificulta el acceso ágil al conocimiento.

Ante este contexto, surge la necesidad de soluciones inteligentes que permitan una interacción más natural y eficiente con la información documental.

Este proyecto propone el desarrollo de un sistema basado en la arquitectura Retrieval-Augmented Generation (RAG), que combina técnicas de recuperación de información con modelos avanzados de lenguaje natural. Gracias a este enfoque, los usuarios podrán cargar documentos y realizar preguntas en lenguaje natural, obteniendo respuestas precisas fundamentadas en fragmentos relevantes del texto original, incluyendo referencias específicas como páginas o secciones.

Adicionalmente, el proyecto analizará distintas estrategias para optimizar la recuperación de información, evaluando diferentes métodos de segmentación documental, variaciones en el número de resultados recuperados (top-k) y la aplicación de técnicas de reranking.

Como complemento orientado al aprendizaje, se incorporará un "modo examen" que permitirá generar automáticamente preguntas basadas en el contenido del documento, facilitando la autoevaluación de los usuarios.

### 3.1 Objetivo General

Desarrollar un sistema inteligente basado en la arquitectura RAG que permita consultar documentos digitales mediante lenguaje natural, facilitando la localización precisa de información y mejorando la eficiencia en el acceso al conocimiento organizacional.

### 3.2 Objetivos Específicos

- Diseñar un sistema que permita la carga y procesamiento automático de documentos digitales.
- Implementar mecanismos de segmentación del contenido para facilitar su indexación y recuperación.
- Desarrollar un sistema de consultas en lenguaje natural que genere respuestas basadas en fragmentos relevantes del documento.
- Incorporar referencias específicas (página, sección) que respalden la veracidad de las respuestas.
- Analizar el impacto de distintas estrategias de segmentación en la calidad de la recuperación de información.
- Evaluar la influencia del número de resultados recuperados (top-k) en la precisión de las respuestas.
- Investigar el uso de técnicas de reranking para mejorar la relevancia de los resultados obtenidos.
- Implementar un modo examen capaz de generar preguntas automáticas a partir del contenido del documento.

### 3.3 Metodología

El desarrollo del proyecto se llevará a cabo mediante una metodología aplicada orientada al diseño e implementación de una solución tecnológica basada en inteligencia artificial. El proceso se dividirá en fases progresivas que permitirán abordar el problema de forma estructurada.

#### 3.3.1 Identificación y Análisis del Problema

La problemática principal identificada en las organizaciones es la dificultad para acceder de forma rápida y precisa a información relevante contenida en documentos digitales extensos. Aunque las herramientas de búsqueda tradicionales permiten localizar palabras clave, no son capaces de interpretar el contexto semántico de las consultas ni de ofrecer respuestas directas fundamentadas en el contenido documental. Como consecuencia:

- Se incrementa el tiempo necesario para encontrar información relevante.
- Se reduce la productividad de los empleados.
- Se dificulta la reutilización del conocimiento corporativo.
- Se genera frustración en los usuarios al no encontrar respuestas precisas.

Este análisis justifica la necesidad de implementar sistemas inteligentes capaces de comprender el lenguaje natural y relacionarlo con el contenido semántico de los documentos.

### 3.3.2 Planificación de la Solución

Para resolver la problemática identificada, se propone el diseño de una aplicación basada en la arquitectura RAG compuesta por los siguientes módulos:

#### 1. Ingesta de documentos

Extracción y limpieza automática del texto contenido en archivos PDF y otros formatos compatibles.

#### 2. Segmentación del contenido (Chunking)

División del texto en fragmentos manejables mediante diferentes estrategias:

- Segmentación por tamaño fijo
- Segmentación por párrafos
- Segmentación basada en estructura (títulos y secciones)

#### 3. Indexación semántica

Conversión de los fragmentos en representaciones vectoriales (embeddings) e indexación en una base de datos vectorial que permita búsquedas eficientes por similitud.

#### 4. Recuperación de información

Obtención de los fragmentos más relevantes en función de la consulta del usuario, evaluando distintas configuraciones del parámetro top-k.

#### 5. Reranking de resultados

Aplicación de técnicas de reordenación para priorizar los fragmentos más relevantes y mejorar la calidad de la respuesta final.

#### 6. Generación de respuestas

Uso de un modelo de lenguaje que genere respuestas en lenguaje natural apoyándose exclusivamente en los fragmentos recuperados, incluyendo referencias documentales.

## 7. Modo examen

Generación automática de preguntas a partir del contenido documental para facilitar la autoevaluación del usuario.

## 4. Identificación y descripción de un problema al interior de la organización

En las organizaciones actuales, la gestión del conocimiento constituye un elemento estratégico para la toma de decisiones, la eficiencia operativa y la innovación. Gran parte de este conocimiento se encuentra almacenado en documentos digitales como manuales técnicos, informes internos, normativas, procedimientos y documentación de proyectos.

Sin embargo, a pesar de la digitalización de la información, muchas organizaciones carecen de herramientas avanzadas que permitan acceder de forma rápida y precisa a contenidos específicos dentro de documentos extensos. Esta limitación genera ineficiencias operativas, pérdida de tiempo y dificultades en la reutilización del conocimiento corporativo.

El problema se vuelve especialmente relevante en entornos donde los empleados necesitan consultar información técnica o normativa de forma frecuente, ya que los métodos tradicionales de búsqueda requieren revisar manualmente grandes volúmenes de información o realizar búsquedas por palabras clave que no siempre devuelven resultados precisos ni contextualizados.

En este contexto, se identifica la necesidad de intervenir en los procesos de gestión documental mediante soluciones tecnológicas basadas en inteligencia artificial que permitan optimizar el acceso a la información organizacional.

### 4.1 Descripción de la empresa a intervenir

La organización objeto de intervención es una empresa de carácter tecnológico orientada a la gestión de información y conocimiento digital. Su actividad principal se centra en la elaboración, almacenamiento y consulta de documentación técnica, informes operativos y manuales de procedimiento que sirven de apoyo a distintos departamentos internos.

La empresa dispone de un repositorio documental amplio compuesto principalmente por archivos en formato PDF, utilizados como soporte para la transmisión de conocimiento entre áreas técnicas, administrativas y de gestión.

El crecimiento continuo del volumen de documentación ha generado la necesidad de implementar herramientas más eficientes que faciliten la consulta de información relevante y mejoren los tiempos de respuesta en los procesos internos.

#### 4.1.1 Misión

La misión de la organización es gestionar y proporcionar información técnica y operativa de manera eficiente, garantizando que el conocimiento corporativo esté disponible para los empleados cuando lo necesiten, facilitando la toma de decisiones y la mejora continua de los procesos internos.

#### 4.1.2 Visión

La visión de la empresa es convertirse en una organización altamente digitalizada e innovadora en la gestión del conocimiento, incorporando tecnologías avanzadas de inteligencia artificial que optimicen el acceso a la información y mejoren la productividad de sus equipos de trabajo.

#### 4.1.3 Mapa de procesos

El funcionamiento de la organización se estructura en tres grandes tipos de procesos:

##### Procesos estratégicos

- Planificación organizacional
- Gestión de la innovación
- Dirección tecnológica

##### Procesos operativos

- Elaboración de documentación técnica
- Gestión documental
- Consulta y recuperación de información
- Soporte a departamentos internos

##### Procesos de apoyo

- Gestión de sistemas informáticos
- Administración interna
- Formación de personal

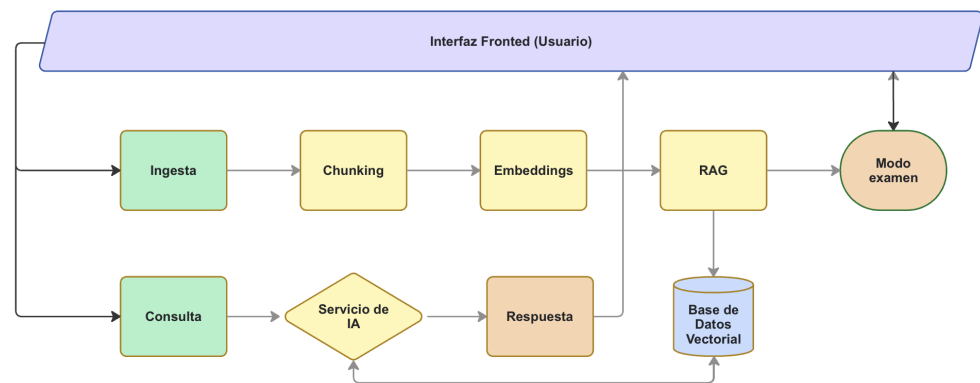
##### Mapa específico de proceso

Aunque el mapa de procesos organizacional permite entender la estructura general de la empresa, resulta necesario complementar esta visión con un mapa específico del proceso intervenido.

## ■ Sistema de Consulta Inteligente sobre Documentos mediante RAG

En este sentido, se ha diseñado un mapa detallado del flujo de gestión documental y recuperación de información, incorporando la solución propuesta basada en arquitectura RAG, lo que permite visualizar de forma clara la mejora introducida en el proceso.

- Ingesta
- Chunking
- Embeddings
- RAG
- Consulta
- Respuesta
- Modo examen



### 4.1.4 Descripción del proceso a intervenir

El proceso de gestión documental comprende las actividades relacionadas con el almacenamiento, organización y consulta de documentos digitales utilizados por los empleados de la organización.

Actualmente, cuando un trabajador necesita información específica contenida en un documento extenso, debe:

1. Localizar el archivo dentro del repositorio documental
2. Abrir el documento manualmente.
3. Revisar el contenido completo o realizar búsquedas por palabras clave.
4. Interpretar la información encontrada para determinar si responde a su necesidad.

Este procedimiento implica un consumo elevado de tiempo y esfuerzo, especialmente cuando los documentos contienen gran cantidad de información técnica o normativa. Además, las búsquedas por palabras clave no permiten comprender el contexto semántico de la consulta, lo que dificulta encontrar respuestas precisas cuando la información está redactada con terminología diferente.

#### 4.1.5 Descripción del problema que presenta el proceso

El proceso actual de recuperación de información presenta múltiples deficiencias que afectan a la eficiencia operativa de la organización:

##### Ineficiencia temporal

Los empleados invierten un tiempo considerable en revisar documentos extensos para localizar información específica.

##### Baja precisión en las búsquedas

Las herramientas tradicionales de búsqueda textual no interpretan el significado de las consultas, limitándose a coincidencias literales de palabras clave.

##### Dificultad en la reutilización del conocimiento

El conocimiento existente en la organización no se aprovecha de forma óptima debido a la complejidad para acceder a información relevante.

##### Sobrecarga operativa del personal

La necesidad de revisar manualmente grandes volúmenes de documentación genera fatiga y reduce la productividad.

##### Falta de herramientas inteligentes

La organización no dispone de sistemas basados en inteligencia artificial que permitan realizar consultas en lenguaje natural ni obtener respuestas directas fundamentadas en el contenido documental.

Como consecuencia, el proceso actual limita la eficiencia organizacional, ralentiza la toma de decisiones y dificulta la gestión efectiva del conocimiento corporativo.

Esta situación justifica la necesidad de desarrollar e implementar una solución tecnológica basada en inteligencia artificial que optimice la recuperación de información documental.

### 5. Establecimiento de planes de acción para la solución del problema

Con el objetivo de resolver las deficiencias identificadas en el proceso de gestión documental de la organización, se establece un plan de acción estructurado que permite diseñar, desarrollar e implementar una solución tecnológica basada en inteligencia artificial.

El plan se organiza en fases progresivas que permiten construir el sistema de forma incremental, validar su funcionamiento y evaluar su impacto en la mejora del acceso a la información.

## 5.1 Establecimiento de objetivos

### Objetivo general

Implementar una aplicación inteligente basada en la arquitectura Retrieval-Augmented Generation (RAG) que optimice la recuperación de información contenida en documentos digitales y facilite la interacción mediante lenguaje natural.

### Objetivos específicos

- Diseñar una aplicación funcional que permita la carga y procesamiento automático de documentos digitales.
- Implementar un sistema de indexación semántica que permita búsquedas eficientes por similitud textual.
- Desarrollar un sistema de consultas en lenguaje natural que proporcione respuestas fundamentadas en el contenido documental.
- Incorporar mecanismos que mejoren la precisión de los resultados, como técnicas de reranking.
- Evaluar distintas estrategias de segmentación de documentos para determinar su impacto en la calidad de la recuperación de información.
- Implementar un modo examen que permita la generación automática de preguntas para reforzar la comprensión del contenido.
- Medir el rendimiento del sistema mediante métricas objetivas de precisión, relevancia y experiencia de usuario.

## 5.2 Descripción de la solución

La solución propuesta consiste en el desarrollo de una aplicación interactiva que integra técnicas de recuperación de información y modelos de lenguaje natural mediante la arquitectura RAG.

El sistema permitirá a los usuarios cargar documentos en formato PDF o Word y realizar consultas en lenguaje natural sobre su contenido. A partir de estas consultas, la aplicación recuperará fragmentos relevantes del documento utilizando búsqueda semántica y generará respuestas fundamentadas en dichos fragmentos, incluyendo referencias específicas como páginas o secciones.

Desde el punto de vista funcional, la solución incorpora cuatro componentes principales:

### Interfaz de usuario

Aplicación web desarrollada con Streamlit que permite la carga de documentos, la visualización del estado de procesamiento y la interacción mediante preguntas en lenguaje natural.

## ■ Sistema de Consulta Inteligente sobre Documentos mediante RAG

### Procesamiento documental

Módulo encargado de extraer, limpiar y estructurar el texto de los documentos, organizándolo por páginas y secciones para facilitar su posterior análisis.

### Motor de recuperación de información (RAG)

Sistema que segmenta el contenido en fragmentos (chunking), genera representaciones vectoriales (embeddings) y almacena la información en una base de datos vectorial para permitir búsquedas por similitud semántica.

### Funcionalidades avanzadas

Incluyen técnicas de reranking para mejorar la relevancia de los resultados, generación de citas precisas y un modo examen que permite crear preguntas automáticas para evaluar la comprensión del usuario.

## 5.3 Secuencia de actividades para la ejecución de la solución

La implementación del proyecto se organiza en un plan de trabajo de cuatro semanas, estructurado en fases progresivas de desarrollo:

### Fase 1: Base de la aplicación e ingestión documental

#### Objetivo:

Construir la estructura inicial del sistema y el flujo de procesamiento de documentos.

#### Actividades principales:

- Configuración del repositorio del proyecto, entorno de desarrollo y dependencias.
- Desarrollo de una interfaz básica en Streamlit para la carga de documentos.
- Implementación de la extracción de texto desde archivos PDF y Word.
- Aplicación de procesos de limpieza básica del texto.
- Definición de un formato interno de organización documental por secciones y páginas.

#### Entregables:

Aplicación funcional con interfaz básica y sistema de extracción de texto validado con documentos de prueba.

### Fase 2: Indexado semántico y recuperación de información (RAG básico)

#### Objetivo:

## ■ Sistema de Consulta Inteligente sobre Documentos mediante RAG

Implementar el núcleo del sistema de recuperación de información.

### Actividades principales:

- Implementación de estrategias de segmentación documental (chunking):
  - Fragmentación por tamaño fijo
  - Fragmentación por párrafos
  - Fragmentación por títulos o secciones
- Desarrollo de un módulo de generación de embeddings desacoplado del proveedor del modelo.
- Integración de una base de datos vectorial para almacenamiento de representaciones semánticas.
- Implementación de búsqueda por similitud mediante parámetro top-k.
- Generación de respuestas iniciales mostrando fragmentos relevantes recuperados.

### Entregables:

Sistema RAG funcional que permita cargar documentos, indexarlos y recuperar información relevante con evidencias textuales.

### Fase 3: Mejora de calidad: reranking, citas y evaluación

#### Objetivo:

Optimizar la precisión de las respuestas y evaluar el rendimiento del sistema.

#### Actividades principales:

- Incorporación de técnicas de reranking para priorizar resultados más relevantes.  
Mejora del sistema de citas mediante referencias precisas de página y sección.  
Resaltado de fragmentos relevantes dentro del texto original.  
Creación de un conjunto de datos de evaluación compuesto por preguntas asociadas a documentos.  
Definición de métricas de evaluación:
  - Exactitud de las respuestas
  - Relevancia de los fragmentos recuperados
  - Tiempo de respuesta
  - Experiencia de usuario

### Entregables:

Sistema de preguntas y respuestas con citas verificables y conjunto de herramientas de evaluación con resultados preliminares.

### Fase 4: Implementación del modo examen y cierre del proyecto

#### Objetivo:

## ■ Sistema de Consulta Inteligente sobre Documentos mediante RAG

Incorporar funcionalidades educativas y consolidar la solución final.

### Actividades principales:

- Desarrollo del modo examen con generación automática de preguntas:
  - Preguntas tipo test
  - Preguntas abiertas
  - Tarjetas de estudio (flashcards)
- Implementación de un sistema de validación de respuestas
- Incorporación de historial de sesiones de usuario.
- Desarrollo de opciones de exportación de resultados en formatos CSV y JSON.
- Optimización de experiencia de usuario, gestión de errores y registro de eventos.
- Elaboración de documentación técnica y análisis comparativo de resultados.

### Entregables:

Aplicación completa con sistema de consulta documental, modo examen y documentación técnica final.

## 5.4 Descripción de recursos necesarios para la ejecución de la solución

### Recursos tecnológicos

- Lenguaje de programación Python
- Framework de desarrollo web Streamlit
- Librerías de procesamiento documental (PDF y Word)
- Modelos de generación de *embeddings*
- Base de datos vectorial para almacenamiento semántico
- Entornos de desarrollo integrados
- Infraestructura informática para pruebas y ejecución del sistema

### Recursos humanos

- Desarrollador principal del proyecto
- Tutor académico para supervisión metodológica
- Usuarios de prueba para evaluación del sistema

### Recursos temporales

El proyecto se desarrollará en un periodo estimado de cuatro semanas, distribuidas en fases de diseño, implementación, validación y documentación.

### Recursos de información

- Documentación técnica sobre arquitectura RAG

## ■ Sistema de Consulta Inteligente sobre Documentos mediante RAG

- Artículos científicos sobre recuperación de información
- Manuales de procesamiento de lenguaje natural
- Repositorios de código y bibliotecas especializadas

### 5.5 Tabla por fases

| Fase                                                | Objetivo                                                                           | Actividades principales                                                                                                                                                                                                                                                                                                                            | Entregables                                                                               |
|-----------------------------------------------------|------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| <b>Fase 1: Base de la aplicación e ingestión</b>    | Construir la estructura inicial del sistema y el flujo de procesamiento documental | <ul style="list-style-type: none"> <li>• Configuración del entorno y repositorio</li> <li>• Desarrollo de UI básica en Streamlit</li> <li>• Carga de documentos PDF/Word</li> <li>• Extracción y limpieza de texto</li> <li>• Organización interna por páginas y secciones</li> </ul>                                                              | Aplicación básica funcional con extracción de texto validada                              |
| <b>Fase 2: Indexado y recuperación (RAG básico)</b> | Implementar el núcleo del sistema de recuperación semántica                        | <ul style="list-style-type: none"> <li>• Segmentación documental (tamaño fijo, párrafos, secciones)</li> <li>• Generación de embeddings</li> <li>• Integración de base de datos vectorial</li> <li>• Búsqueda por similitud (top-k)</li> <li>• Visualización de fragmentos relevantes</li> </ul>                                                   | Sistema RAG funcional con evidencias textuales recuperadas                                |
| <b>Fase 3: Mejora de calidad y evaluación</b>       | Optimizar la precisión de respuestas y medir rendimiento                           | <ul style="list-style-type: none"> <li>• Implementación de reranking</li> <li>• Mejora de citas (página/sección)</li> <li>• Resaltado de fragmentos relevantes</li> <li>• Creación de dataset de evaluación</li> <li>• Definición de métricas (exactitud, relevancia, tiempo, UX)</li> </ul>                                                       | Sistema de preguntas-respuestas con citas verificables + informe preliminar de evaluación |
| <b>Fase 4: Modo examen y cierre</b>                 | Incorporar funcionalidades educativas y consolidar la solución final               | <ul style="list-style-type: none"> <li>• Generación automática de preguntas (test, abiertas, flashcards)</li> <li>• Validación de respuestas con rúbrica</li> <li>• Historial de sesiones</li> <li>• Exportación de resultados (CSV/JSON)</li> <li>• Mejoras UX, logs y manejo de errores</li> <li>• Documentación final y conclusiones</li> </ul> | Aplicación completa operativa + documentación técnica + conclusiones finales              |

## 6. Definición del alcance de la solución, partes interesadas, cronograma y presupuesto

### 6.1 Alcance de la solución

El alcance de la solución propuesta comprende el diseño, desarrollo e implementación de un sistema inteligente basado en la arquitectura Retrieval-Augmented Generation (RAG), orientado a optimizar la consulta de documentos digitales mediante lenguaje natural.

#### Objetivo del alcance

Desarrollar una aplicación funcional que permita mejorar la eficiencia en la recuperación de información, reduciendo los tiempos de búsqueda y aumentando la precisión de los resultados.

#### Resultados esperados

- Implementación de un sistema capaz de procesar documentos PDF y Word.
- Generación de respuestas en lenguaje natural basadas en contenido documental.
- Mejora en la precisión mediante técnicas de embeddings y reranking.
- Incorporación de un modo examen para la evaluación del conocimiento.
- Validación del sistema mediante métricas de rendimiento y experiencia de usuario.

#### Límites del alcance

- No se contempla integración con sistemas empresariales externos (ERP, CRM).
- No incluye despliegue en entornos productivos a gran escala.
- El sistema se desarrolla en un entorno de prueba funcional.
- No se implementa soporte multilingüe avanzado.

### 6.2 Identificación de responsables de las actividades para ejecución

Para garantizar una ejecución eficiente del proyecto, se identifican las siguientes partes interesadas:

**Tabla 1. Roles y responsabilidades**

| Rol                     | Responsabilidades principales                   | Participación         |
|-------------------------|-------------------------------------------------|-----------------------|
| Desarrollador principal | Diseño, desarrollo e implementación del sistema | Ejecución técnica     |
| Tutor académico         | Supervisión metodológica y validación           | Control y seguimiento |

## ■ Sistema de Consulta Inteligente sobre Documentos mediante RAG

|                          |                                             |                |
|--------------------------|---------------------------------------------|----------------|
| Usuarios de prueba       | Evaluación funcional del sistema            | Validación     |
| Responsable del proyecto | Planificación, coordinación y documentación | Gestión global |

Esta estructura organizativa permite una adecuada distribución de responsabilidades y asegura la correcta coordinación del proyecto.

### 6.3 Cronograma de ejecución

El proyecto se desarrollará en un periodo estimado de cuatro semanas, organizadas en fases progresivas.

#### Hitos clave

- Sistema base implementado (Semana 1)
- Sistema RAG funcional (Semana 2)
- Sistema optimizado y evaluado (Semana 3)
- Proyecto finalizado (Semana 4)

Este cronograma permite realizar un seguimiento continuo del progreso del proyecto y facilita la evaluación de cada fase.

**Tabla 2. Cronograma de actividades**

| Semana   | Fecha inicio | Fecha fin | Fase                     | Actividades principales                               | Hito                   |
|----------|--------------|-----------|--------------------------|-------------------------------------------------------|------------------------|
| Semana 1 | 01/03/26     | 07/03/26  | Fase 1: Base del sistema | Configuración entorno, interfaz, ingestión documental | Sistema base funcional |
| Semana 2 | 08/03/26     | 14/03/26  | Fase 2: RAG básico       | Chunking, embeddings, base vectorial                  | Sistema RAG operativo  |
| Semana 3 | 15/03/26     | 21/03/26  | Fase 3: Mejora           | Reranking, citas, evaluación                          | Sistema optimizado     |
| Semana 4 | 22/03/26     | 28/03/26  | Fase 4: Cierre           | Modo examen, mejoras UX, documentación                | Entrega final          |

## 6.4 Presupuesto

El presupuesto del proyecto se ha estimado considerando los recursos necesarios para su desarrollo en un entorno académico.

### Justificación del presupuesto

El principal coste del proyecto corresponde al desarrollo del sistema, que requiere dedicación técnica especializada. La supervisión académica garantiza la calidad metodológica del trabajo. Los costes tecnológicos se mantienen bajos gracias al uso de herramientas open source y al uso controlado de APIs. Los costes indirectos representan los recursos básicos necesarios para la ejecución del proyecto.

**Tabla 3. Presupuesto**

| Rubro                                      | Valor   |
|--------------------------------------------|---------|
| Desarrollo del sistema (4 semanas)         | 2.000 € |
| Supervisión académica                      | 500 €   |
| Uso de APIs / modelos de IA                | 100 €   |
| Costes indirectos (electricidad, internet) | 50 €    |
| Total                                      | 2.650 € |

## 6.5 Evaluación y seguimiento

Para garantizar una implementación eficiente, se establecen mecanismos de evaluación continua:

- Seguimiento semanal del avance del proyecto.
- Validación de entregables en cada fase.
- Evaluación mediante métricas (precisión, relevancia y tiempo de respuesta).
- Recogida de feedback de usuarios de prueba.

Este enfoque permite detectar posibles desviaciones, aplicar mejoras durante el desarrollo y asegurar el cumplimiento de los objetivos definidos.

## 7. Referencias bibliográficas

1. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-augmented generation for knowledge-intensive NLP tasks*. <https://arxiv.org/abs/2005.11401>
2. Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). <https://web.stanford.edu/~jurafsky/slp3/>
3. OpenAI. (2024). *Documentation on language models and embeddings*. <https://platform.openai.com/docs>
4. Pinecone Systems. (2024). *Vector databases explained*. <https://www.pinecone.io/learn/vector-database/>
5. LangChain. (2024). *Building applications with LLMs*. <https://docs.langchain.com/>
6. Streamlit. (2024). *Web app framework for data science*. <https://docs.streamlit.io/>