

ESTRUCTURA DE TRABAJO: PROYECTO INTEGRADOR

Sistema RAG Multimodal para Estimación de Cantidades de Obra en Construcción

Institución: Tecnológico de Monterrey - Maestría en Inteligencia Artificial Aplicada

Sponsor: ConstructAI

Duración: 10 semanas

Mentor: Fernando Maytorena (CTO & Co-founder, ConstructAI)

OBJETIVOS DEL PROYECTO

Objetivo Principal

Desarrollar el **primer sistema RAG multimodal en América Latina** específico para la industria de construcción, que automatice la creación y consulta de knowledge bases para documentos técnicos, eliminando bottlenecks humanos y mejorando la precisión en extracción contextual.

Objetivos Específicos

1. **Desarrollar sistema RAG multimodal avanzado** con hybrid search y reranking multimodal
 2. **Diseñar sistema de detección de objetos** automatizado específicamente para alimentar el RAG
 3. **Crear pipeline end-to-end** desde documentos PDF hasta retrieval contextual preciso
-

FILOSOFÍA DE DESARROLLO

Enfoque Requirements-Driven

El proyecto sigue una **lógica inversa intencional**:

1. **Primero RAG:** Descubrir limitaciones y requisitos específicos del retrieval multimodal
2. **Después Segmentación:** Diseñar detección de objetos específicamente para optimizar inputs del RAG

3. **Finalmente Integración:** Crear pipeline cohesivo basado en aprendizajes de ambas etapas

Prioridades Técnicas

- **Prioridad #1:** Calidad y robustez del sistema RAG multimodal
 - **Prioridad #2:** Sistema de detección automática que supere segmentación manual actual
 - **Prioridad #3:** Evaluación rigurosa y benchmarking contra baselines
-

METODOLOGÍA DE TRABAJO

Estructura de Entregables

Cada semana incluye:

- **Entregable técnico concreto** (código, sistema, análisis)
- **Investigación guiada** con criterios específicos para autodescubrimiento
- **Evaluación cuantitativa** con métricas definidas
- **Documentación técnica** de decisiones y resultados

Patrón de Instrucciones

Todas las semanas siguen el patrón de **"investigación guiada"**:

"Investiguen y experimenten con [técnicas específicas], priorizando aquellos enfoques que [criterios técnicos]. Justifiquen su selección basándose en [evidencia requerida]. Evalúen usando [métricas específicas]."

ESTRUCTURA SEMANAL DETALLADA

FASE 1: SISTEMA RAG MULTIMODAL (Semanas 1-4)

Prioridad: Desarrollar RAG robusto que supere baseline textual existente

SEMANA 1: RAG Básico + Baseline

Objetivo: Establecer sistema RAG funcional usando knowledge base textual existente

Entregable: Sistema RAG baseline completamente funcional con evaluación inicial

Instrucción:

Implementen un sistema RAG básico utilizando el knowledge base textual existente (archivos markdown). Investiguen y seleccionen un modelo de embedding multimodal de uso general (ej: Gemini, OpenAI), implementen estrategia de chunking básica, y configuren un retrieval system funcional. Este sistema servirá como baseline para todas las comparaciones futuras. Evalúen el sistema con 5 queries representativas seleccionadas por el mentor y documenten métricas de performance (Precision@5).

Criterios de Evaluación:

- ☐ Sistema RAG funcional end-to-end
- ☐ Evaluación con 5 queries seleccionadas por mentor
- ☐ Métricas baseline documentadas (Precision@5)
- ☐ Justificación técnica de decisiones de embedding y chunking
- ☐ Código documentado y reproducible

Investigación Requerida:

- Modelos de embedding multimodales mainstream (Gemini, OpenAI)
- Estrategias de chunking para documentos largos
- Métricas de evaluación para sistemas RAG

SEMANA 2: RAG Multimodal + Hybrid Search

Objetivo: Incorporar imágenes segmentadas y comparar semantic vs hybrid search

Entregable: Sistema RAG multimodal con análisis comparativo de estrategias de búsqueda

Instrucción:

Extiendan el sistema RAG baseline para incorporar las imágenes segmentadas existentes. Investiguen modelos de embedding multimodales que procesen simultáneamente texto e imágenes, priorizando aquellos optimizados para documentos técnicos y tareas de retrieval. Implementen y comparen 2 estrategias de búsqueda: (1) Semantic search (búsqueda semántica usando embeddings multimodales), (2) Hybrid search (combinación de semantic search + keyword search). Evalúen performance de cada estrategia usando 5 queries seleccionadas por el mentor. Analicen en qué tipos de queries cada estrategia es superior y justifiquen técnicamente sus hallazgos.

Criterios de Evaluación:

- ☐ Sistema RAG multimodal funcional (texto + imágenes)
- ☐ Implementación de semantic search y hybrid search
- ☐ Análisis comparativo detallado con métricas cuantitativas
- ☐ Justificación de cuándo usar cada estrategia de búsqueda
- ☐ Mejora medible vs baseline textual de Semana 1 (evaluado con 5 queries por mentor)

Investigación Requerida:

- Modelos de embedding multimodales (ej. ColPali, Gemini-embedding, BGE-M3)
- Hybrid search techniques (combinación semantic + keyword)
- Trade-offs entre semantic search vs hybrid search en dominio técnico

SEMANA 3: Reranking Multimodal + Query Processing

Objetivo: Optimizar pipeline con reranking avanzado y procesamiento de queries

Entregable: Pipeline retriever→reranker optimizado + sistema de query expansion

Instrucción:

Implementen reranking multimodal para reordenar resultados de su mejor estrategia de búsqueda de la Semana 2. Investiguen modelos de cross-encoding específicos para documentos técnicos multimodales, priorizando aquellos que puedan manejar texto + imagen simultáneamente. Adicionalmente, implementen al menos 2 estrategias diferentes de query expansion/rewriting específicamente diseñadas para consultas técnicas de construcción (ej: expandir términos técnicos, reformular queries mal estructuradas). Evalúen el impacto incremental de cada componente: (1) búsqueda sola, (2) búsqueda + reranker, (3) búsqueda + reranker + query processing, midiendo mejoras en Precision@5 en cada etapa.

Criterios de Evaluación:

- ☐ Reranking multimodal implementado y funcional
- ☐ Al menos 2 estrategias de query processing diferentes
- ☐ Análisis incremental de mejora por componente (búsqueda → +reranker → +query processing)
- ☐ Evaluación con 5 queries seleccionadas por mentor
- ☐ Optimización medible del pipeline completo

Investigación Requerida:

- Modelos de cross-encoding para reranking multimodal
 - Query expansion techniques para dominios técnicos específicos
 - Análisis de impacto incremental por componente en sistemas RAG
-

SEMANA 4: Evaluación Robusta + Análisis de Limitaciones

Objetivo: Evaluación completa con ground truth + especificaciones para segmentación

Entregable: Sistema RAG completo evaluado + recomendaciones detalladas para segmentación

Instrucción:

Evalúen el sistema RAG multimodal completo (hybrid search + reranker + query processing) usando 5 queries seleccionadas por el mentor. Realicen análisis detallado de failure cases en estas queries, identificando exactamente qué tipos de contenido visual/textual el sistema no puede recuperar efectivamente. Documenten patrones en errores y limitaciones específicas del sistema actual. Basándose en estos hallazgos, generen especificaciones técnicas detalladas sobre qué características específicas debería tener un sistema de segmentación automática para maximizar el performance del RAG (tipos de objetos prioritarios, calidad de recortes, metadata requerido, etc.).

Criterios de Evaluación:

- ☐ Evaluación completa del sistema RAG con 5 queries seleccionadas por mentor
- ☐ Análisis detallado de failure cases con categorización
- ☐ Especificaciones técnicas para sistema de segmentación ideal
- ☐ Benchmarking final vs todos los baselines anteriores
- ☐ Recomendaciones específicas basadas en evidencia

Investigación Requerida:

- Metodologías de evaluación robusta para sistemas RAG
 - Análisis de failure modes en retrieval multimodal
 - Requirements engineering para sistemas de computer vision
-

FASE 2: DETECCIÓN Y CLASIFICACIÓN DE OBJETOS (Semanas 5-8)

Prioridad: Diseñar segmentación automática optimizada para inputs del RAG

SEMANA 5: Análisis de Requirements + Arquitectura de Detección

Objetivo: [PLACEHOLDER - Definir arquitectura basada en findings de RAG]

Entregable: [PLACEHOLDER - Sistema de detección básico]

Instrucción: [PLACEHOLDER - Implementar detección automática basada en requirements del RAG]

SEMANA 6: Optimización de Detección para RAG

Objetivo: [PLACEHOLDER - Optimizar detección específicamente para calidad de inputs RAG]

Entregable: [PLACEHOLDER - Sistema optimizado para informatividad vs precisión general]

SEMANA 7: Clasificación y Metadata Generation

Objetivo: [PLACEHOLDER - Sistema de clasificación y generación de metadata rich]

Entregable: [PLACEHOLDER - Pipeline completo de segmentación]

SEMANA 8: Evaluación y Validación de Segmentación

Objetivo: [PLACEHOLDER - Validar que segmentación mejora performance del RAG]

Entregable: [PLACEHOLDER - Sistema validado con métricas de impacto en RAG]

FASE 3: INTEGRACIÓN Y EVALUACIÓN FINAL (Semanas 9-10)

Prioridad: Pipeline end-to-end optimizado y evaluación académica rigurosa

SEMANA 9: Integración End-to-End

Objetivo: [PLACEHOLDER - Integrar segmentación automática + RAG multimodal]

Entregable: [PLACEHOLDER - Sistema completo funcionando automáticamente]

SEMANA 10: Evaluación Final y Documentación Académica

Objetivo: Evaluación rigurosa del sistema completo + documentación para publicación

Entregable: Sistema final evaluado con dataset completo + documentación académica

Instrucción: [PLACEHOLDER - Evaluación final con 35-50 queries completas anotadas por mentor + análisis académico riguroso para publicación]



CRITERIOS DE EVALUACIÓN FINAL

Criterios Técnicos (60%)

- **Sistema RAG Multimodal:** Supera baseline textual en $\geq 20\%$ (Precision@5)
- **Pipeline End-to-End:** Funciona completamente sin intervención manual
- **Benchmarking:** Evaluación rigurosa con ground truth de 35-50 queries
- **Robustez:** Sistema maneja edge cases y diferentes tipos de documentos

Criterios de Investigación (25%)

- **Contribución Académica:** Hallazgos novel sobre RAG multimodal en dominio técnico
- **Metodología:** Experimentación rigurosa con justificación técnica sólida
- **Reproducibilidad:** Código documentado y experimentos replicables
- **Análisis Crítico:** Identificación clara de limitaciones y trabajo futuro

Criterios de Impacto (15%)

- **Aplicabilidad Industrial:** Sistema viable para uso en producción
 - **Escalabilidad:** Diseño que soporte crecimiento de dataset
 - **Documentación:** Calidad suficiente para transferencia a equipos de desarrollo
-

Preparado por: Fernando Maytorena, CTO ConstructAI

Institución: Tecnológico de Monterrey

Programa: Maestría en Inteligencia Artificial Aplicada

Fecha: Septiembre 2025