



RAG Multimodal Multiagente

Researching en Diversas fuentes:

> Serval Projects, Udemy, Platzi, ChatGPT y Claude.

Retrieval Design Architecture

2

¿Cuál es el área de la puerta ubicada en la página 12, sección 1?

1. Query Processor Node

Preparar la consulta para el análisis posterior. Este nodo normalizaría el texto y extraería datos clave.

¿Qué está pidiendo exactamente el usuario?

Un dato específico: "área"

De un elemento específico: "puerta"

En una ubicación específica: "página 12, sección 1"

¿Qué tipo de información necesito buscar?

Información numérica (el área)

Información visual (identificar la puerta)

Información espacial (localizar página y sección)

Este nodo recibe la pregunta cruda del usuario, exactamente como la escribió, con posibles errores tipográficos, lenguaje coloquial, o ambigüedades.

El nodo limpia y normaliza la consulta, identifica el idioma, corrige errores obvios y extrae las palabras clave más importantes. Es como cuando un traductor profesional no solo traduce las palabras, sino que entiende la intención detrás de ellas.

Output: {

"original_query": "¿Cuál es el área de la puerta ubicada en la página 12, sección 1?",

"normalized_query": "cual es el area de la puerta ubicada en la pagina 12 seccion 1",

"language": "spanish",

"key_tokens": ["area", "puerta", "pagina 12", "seccion 1"],

"query_intent": "find_specific_measurement",

"query_type_indicators": ["measurement_request",

"spatial_location", "object_identification"]

}

Retrieval Design Architecture

3

¿Cuál es el área de la puerta ubicada en la página 12, sección 1?

1. Query Processor Node

Prompt:

Eres un experto en procesamiento de consultas sobre planos arquitectónicos y de construcción. Tu trabajo es analizar la consulta del usuario y extraer información estructurada.

Dada esta consulta del usuario: '{user_query}' Analiza y extrae:

¿Qué tipo de información específica está pidiendo el usuario? (medida, ubicación, especificación técnica, etc.)+

¿Qué elemento arquitectónico o de construcción menciona? (puerta, ventana, habitación, escalera, etc.)

¿Hay restricciones espaciales mencionadas? (página, sección, planta, área específica) ¿Qué nivel de especificidad tiene la consulta? (general, específica, muy específica)

¿Qué tipo de respuesta espera el usuario? (numérica, descriptiva, visual, listado) Proporciona tu análisis en formato estructurado, identificando claramente cada componente. Si hay ambigüedades en la consulta, señálalas específicamente.

Retrieval Design Architecture

4

¿Cuál es el área de la puerta ubicada en la página 12, sección 1?

2. Query Analysis

Este nodo descompone la consulta en sus componentes fundamentales y analiza las implicaciones de cada elemento. Determina qué tipos de búsqueda serán necesarios y qué nivel de complejidad tendrá la consulta.

Expandimos los datos de entrada a mas detalle.

```
Input: {
  "original_query": "¿Cuál es el área de la puerta ubicada en la página 12, sección 1?",
  "normalized_query": "cual es el area de la puerta ubicada en la pagina 12 seccion 1",
  "language": "spanish",
  "key_tokens": ["area", "puerta", "pagina 12", "seccion 1"],
  "query_intent": "find_specific_measurement",
  "query_type_indicators": ["measurement_request", "spatial_location", "object_identification"]
}
```

```
Output: {
  "original_query": "¿Cuál es el área de la puerta ubicada en la página 12, sección 1?",
  "parsed_entities": {
    "target_object": {
      "primary": "puerta",
      "type": "architectural_element",
      "visual_searchable": true
    },
    "desired_property": {
      "property": "area",
      "data_type": "numerical",
      "typical_units": ["m²", "ft²", "cm²"],
      "search_patterns": ["A =", "área =", "area:", "superficie"]
    },
    "spatial_constraints": {
      "page": 12,
      "section": 1,
      "constraint_type": "exact_location",
      "filter_required": true
    }
  },
  "query_complexity": "medium",
  "required_capabilities": {
    "spatial_filtering": true,
    "visual_identification": true,
    "text_extraction": true,
    "numerical_parsing": true
  },
  "expected_result_type": "single_numerical_value"
}
```

Retrieval Design Architecture

5

¿Cuál es el área de la puerta ubicada en la página 12, sección 1?

2. Query Analysis

El Prompt Específico:

"Eres un analista de sistemas de información arquitectónica. Tu trabajo es determinar qué tipos de búsqueda y procesamiento serán necesarios para responder una consulta específica. Información de la consulta procesada: {processed_query_data}

Basándote en este análisis, determina:

- ¿Se necesitará identificación visual de elementos arquitectónicos? ¿De qué tipo específicamente?
- ¿Se requerirá extracción y análisis de texto/anotaciones?
 - ¿Qué patrones textuales deberías buscar?
- ¿Hay necesidad de filtrado espacial?
- ¿Qué tan preciso debe ser?
- ¿La respuesta requerirá cálculos o conversiones adicionales?
- ¿Qué nivel de confianza será aceptable para cada tipo de búsqueda?
- ¿Cuáles son los posibles puntos de fallo y cómo se pueden mitigar?

Además, clasifica la complejidad de la consulta como: simple (un solo tipo de búsqueda), media (dos tipos coordinados), o alta (múltiples tipos con dependencias complejas). Proporciona recomendaciones específicas sobre qué estrategias de búsqueda serían más efectivas."

Retrieval Design Architecture

6

¿Cuál es el área de la puerta ubicada en la página 12, sección 1?

3. Decision Node

Evalúa todos los requerimientos y toma decisiones sobre la mejor manera de ejecutar la búsqueda. Expandimos los datos de entrada a mas detalle.

Input: Query Analysis Node Output

```
Input: {
  "parsed_entities": {
    "target_object": {"primary": "puerta", "visual_searchable": true},
    "desired_property": {"property": "area", "data_type": "numerical"},
    "spatial_constraints": {"page": 12, "section": 1, "filter_required": true}
  },
  "query_complexity": "medium",
  "required_capabilities": {
    "spatial_filtering": true,
    "visual_identification": true,
    "text_extraction": true,
    "numerical_parsing": true
  }
}
```

3. Decision Node

Se obtiene un plan de acción concreto con parámetros específicos.

```
Output: {
  "selected_strategy": "hybrid_visual_text_with_spatial_priority",
  "execution_plan": {
    "phase_1": {
      "name": "spatial_filtering",
      "action": "filter_by_metadata",
      "parameters": {
        "page_number": 12,
        "section_number": 1
      },
      "expected_reduction": "95%",
      "mongodb_operation": "find",
      "collection": "ImageEmbeddings"
    },
    "phase_2": {
      "name": "visual_object_detection",
      "action": "clip_similarity_search",
      "parameters": {
        "target_concept": "puerta",
        "similarity_threshold": 0.75,
        "max_results": 10
      },
      "mongodb_operation": "vector_search",
      "embedding_field": "image_embedding"
    },
    "phase_3": {
      "name": "contextual_text_search",
      "action": "proximity_text_search",
      "parameters": {
        "search_terms": ["área", "A =", "área", "m²"],
        "proximity_radius": 50,
        "confidence_threshold": 0.70
      },
      "text_fields": ["text", "ocr_bboxes"]
    }
  },
  "success_criteria": {
    "minimum_confidence": 0.80,
    "required_data_types": ["visual_match", "numerical_value"]
  }
}
```

3. Decision Node

Se obtiene un plan de acción concreto con parámetros específicos.

El Prompt Específico:

"Eres un arquitecto de sistemas de búsqueda multimodal especializado en planos de construcción. Tu trabajo es diseñar la estrategia óptima de búsqueda basándote en el análisis técnico proporcionado.

Análisis técnico recibido: {technical_analysis}

Tu decisión debe considerar: Eficiencia computacional:

¿Cuál es el orden óptimo de operaciones para minimizar el procesamiento innecesario?

Probabilidad de éxito: ¿Qué estrategia tiene mayor probabilidad de encontrar la información correcta?

Recursos disponibles: ¿Qué capacidades específicas de MongoDB Vector Search, índices de texto, y filtros espaciales debes usar?

Criterios de confianza: ¿Qué umbrales de confianza establecerás para cada fase?

Decide entre estas estrategias principales:

VISUAL_ONLY: Búsqueda puramente basada en reconocimiento visual

TEXT_ONLY: Búsqueda puramente basada en contenido textual/OCR

SPATIAL_ONLY: Búsqueda basada únicamente en metadatos espaciales

HYBRID_VISUAL_TEXT: Combinación coordinada de búsqueda visual y textual

HYBRID_SPATIAL_SEMANTIC: Combinación de filtrado espacial con búsqueda semántica

Para la estrategia elegida, especifica:

Secuencia exacta de operaciones

Parámetros específicos para cada operación

Criterios de éxito para cada fase

Plan de contingencia si una fase falla"

Retrieval Design Architecture

¿Cuál es el área de la puerta ubicada en la página 12, sección 1?

4. Hybrid Query Execution

Ejecuta cada fase del plan secuencialmente, usando los resultados de cada fase para refinar la siguiente. Coordina múltiples tipos de consultas a MongoDB y combina los resultados de manera inteligente.

```
Input: {
  "selected_strategy": "hybrid_visual_text_with_spatial_priority",
  "execution_plan": {
    "phase_1": {"action": "filter_by_metadata", "parameters": {"page_number": 12,
"section_number": 1}},
    "phase_2": {"action": "clip_similarity_search", "parameters": {"target_concept":
"puerta", "similarity_threshold": 0.75}},
    "phase_3": {"action": "proximity_text_search", "parameters": {"search_terms": ["area",
"A =", "área", "m²"]}}
  },
  "success_criteria": {"minimum_confidence": 0.80}
}
```

Retrieval Design Architecture

4. Hybrid Query Execution

La transformación fundamental aquí es que hemos convertido un plan abstracto en resultados concretos de búsqueda, con métricas detalladas que permiten evaluar la calidad de cada hallazgo.

```
Output: {
  "execution_summary": {
    "total_phases": 3, "successful_phases": 3, "total_execution_time": "1.4_seconds",
    "documents_processed": {"initial": 15847, "after_phase_1": 23, "after_phase_2": 3, "final": 2}
  },
  "search_results": [
    {
      "result_id": "segment_p12_s1_001",
      "phase_1_match": {
        "page": 12,
        "section": 1,
        "metadata_confidence": 1.0
      },
      "phase_2_match": {
        "visual_object": "puerta",
        "clip_similarity_score": 0.89,
        "visual_confidence": 0.89,
        "bounding_box": {"x1": 120, "y1": 450, "x2": 180, "y2": 520}
      },
      "phase_3_match": {
        "extracted_text": "A = 2.1 m²",
        "text_confidence": 0.95,
        "text_location": {"x": 185, "y": 465},
        "proximity_to_visual": 15,
        "pattern_matched": "A = [number] m²"
      },
      "combined_metrics": {
        "overall_confidence": 0.92,
        "completeness_score": 1.0,
        "relevance_score": 0.94
      }
    },
    {
      "result_id": "segment_p12_s1_002",
      "phase_1_match": {"page": 12, "section": 1, "metadata_confidence": 1.0},
      "phase_2_match": {"visual_object": "puerta", "clip_similarity_score": 0.76, "visual_confidence": 0.76},
      "phase_3_match": null,
      "combined_metrics": {"overall_confidence": 0.38, "completeness_score": 0.5}
    }
  ],
  "search_statistics": {
    "total_candidates": 2,
    "high_confidence_results": 1,
    "low_confidence_results": 1
  }
}
```

Retrieval Design Architecture

11

4. Hybrid Query Execution

La transformación fundamental aquí es que hemos convertido un plan abstracto en resultados concretos de búsqueda, con métricas detalladas que permiten evaluar la calidad de cada hallazgo.

El Prompt Específico: "Eres un especialista en ejecución de búsquedas multimodales en bases de datos MongoDB especializadas en contenido arquitectónico. Tu trabajo es ejecutar el plan de búsqueda y evaluar los resultados obtenidos.

Plan de ejecución a seguir: {execution_plan}

Para cada fase del plan, debes:

1. Simular la ejecución de la operación especificada
2. Evaluar la calidad de los resultados obtenidos en esa fase
3. Determinar si los resultados cumplen con los criterios establecidos
4. Decidir si continuar con la siguiente fase o ajustar la estrategia

Al evaluar resultados, considera:

- Confianza en la identificación visual (si aplica)
- Precisión en la extracción textual (si aplica)
- Coherencia espacial entre elementos encontrados
- Completitud de la información requerida

Proporciona un reporte detallado de:

- Resultados de cada fase
- Métricas de confianza para cada resultado
- Recomendaciones sobre cuáles resultados son más confiables
- Identificación de cualquier información faltante o ambigua

Si alguna fase no produce resultados satisfactorios, sugiere ajustes específicos a la estrategia."

Retrieval Design Architecture

¿Cuál es el área de la puerta ubicada en la página 12, sección 1?

Nodo 5: Results Ranking

Este nodo aplica algoritmos de ranking que consideran múltiples factores de calidad simultáneamente. Evalúa no solo la confianza individual de cada componente, sino también la coherencia entre componentes y la completitud de la respuesta.

```
Input: {
  "search_results": [
    {
      "result_id": "segment_p12_s1_001",
      "phase_2_match": {"visual_confidence": 0.89, "clip_similarity_score": 0.89},
      "phase_3_match": {"extracted_text": "A = 2.1 m²", "text_confidence": 0.95,
      "proximity_to_visual": 15},
      "combined_metrics": {"overall_confidence": 0.92, "completeness_score": 1.0}
    },
    {
      "result_id": "segment_p12_s1_002",
      "phase_2_match": {"visual_confidence": 0.76},
      "phase_3_match": null,
      "combined_metrics": {"overall_confidence": 0.38, "completeness_score": 0.5}
    }
  ]
}
```

```
Output: {
  "ranked_results": [
    {
      "final_rank": 1,
      "result_id": "segment_p12_s1_001",
      "quality_assessment": {
        "overall_score": 94.2,
        "visual_quality": 89.0,
        "textual_quality": 95.0,
        "spatial_coherence": 98.5,
        "completeness": 100.0
      },
      "extracted_answer": {
        "value": "2.1",
        "unit": "m²",
        "full_text": "A = 2.1 m²",
        "confidence": 0.95
      },
      "supporting_evidence": {
        "visual_element_location": {"x1": 120, "y1": 450, "x2": 180, "y2": 520},
        "text_element_location": {"x": 185, "y": 465},
        "source_segment": "segments/page12_section1_001.png",
        "spatial_relationship": "text_adjacent_to_visual"
      },
      "reliability_indicators": {
        "answer_confidence": "high",
        "evidence_strength": "strong",
        "spatial_validation": "confirmed"
      }
    }
  ],
  "ranking_summary": {
    "total_evaluated": 2,
    "high_quality_results": 1,
    "recommended_result": "segment_p12_s1_001",
    "confidence_in_recommendation": 0.942
  }
}
```

Retrieval Design Architecture

13

¿Cuál es el área de la puerta ubicada en la página 12, sección 1?

Nodo 5: Results Ranking

El Prompt Específico: "Eres un evaluador experto en calidad de resultados para sistemas de información arquitectónica. Tu especialidad es determinar qué tan confiables y útiles son los resultados de búsquedas multimodales.

Resultados de búsqueda recibidos: {search_results} Consulta original del usuario: {original_query}

Para cada resultado, evalúa:

- 1.Consistencia interna: ¿Los componentes visuales y textuales del resultado son coherentes entre sí?
- 2.Relevancia directa: ¿Qué tan directamente responde este resultado a la pregunta específica del usuario?
- 3.Confianza técnica: ¿Qué tan altos son los scores de confianza de cada componente?
- 4.Completitud: ¿El resultado proporciona toda la información que el usuario necesita?
- 5.Contexto espacial: ¿La ubicación del resultado coincide con las especificaciones del usuario?

Criterios específicos para planos arquitectónicos:

- ¿Las medidas encontradas son realistas para el tipo de elemento arquitectónico?
- ¿La proximidad entre elementos visuales y textual sugiere una relación real?
- ¿Los patrones textuales coinciden con notaciones estándar de arquitectura?

Proporciona:

- Ranking de resultados del mejor al peor
- Justificación detallada para el ranking
- Nivel de confianza en la recomendación principal
- Identificación de cualquier inconsistencia o señal de alerta
- Recomendación sobre si los resultados son suficientes para responder al usuario o si se necesita búsqueda adicional"

Retrieval Design Architecture

¿Cuál es el área de la puerta ubicada en la página 12, sección 1?

6: Response Generation

Este nodo toma todos los datos técnicos y los convierte en una respuesta comprensible y útil para el usuario final.

```
Input: {
  "ranked_results": [{
    "final_rank": 1,
    "extracted_answer": {"value": "2.1", "unit": "m²", "confidence": 0.95},
    "quality_assessment": {"overall_score": 94.2},
    "supporting_evidence": {
      "source_segment": "segments/page12_section1_001.png",
      "spatial_relationship": "text_adjacent_to_visual"
    },
    "reliability_indicators": {"answer_confidence": "high", "evidence_strength": "strong"}
  }],
  "original_query": "¿Cuál es el área de la puerta ubicada en la página 12, sección 1?"
}
```

```
Output: {
  "primary_response": "El área de la puerta ubicada en la página 12, sección 1 es de 2.1 metros cuadrados.",
  "detailed_response": "He localizado la puerta específica que buscas en la página 12, sección 1 del plano. El área de esta puerta es de 2.1 metros cuadrados. Esta información se obtuvo mediante la identificación visual del elemento arquitectónico y la extracción de la anotación textual asociada que especifica la medida.",
  "confidence_communication": "Tengo un alto nivel de confianza en esta respuesta (94.2% de precisión) porque tanto la identificación visual como la extracción de texto tuvieron resultados excelentes.",
  "evidence_available": {
    "visual_evidence": "Ubicación exacta del elemento identificado en el plano",
    "textual_evidence": "Anotación original que especifica 'A = 2.1 m²'",
    "spatial_evidence": "Confirmación de ubicación en página y sección especificadas"
  },
  "additional_context": "La medida se encontró como una anotación textual ubicada espacialmente muy cerca del elemento visual identificado, lo que confirma la relación entre ambos componentes.",
  "response_metadata": {
    "query_processing_time": "1.4 seconds",
    "result_confidence": "high",
    "evidence_strength": "strong"
  }
}
```

Flowchart

- ▶ <https://www.mermaidchart.com/app/projects/fd29a72d-dbb5-46ff-9057-8d99a0d9af7f/diagrams/a679c049-564d-4c53-b208-4b064552a2a6/share/invite/eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJkb2N1bWVudEEljoiYTY3OWMwNDktNTY0ZC00YzUzLWlyMDgtNGlwNjQ1NTJhMmE2liwiYWJjZXNzIjoiRWRpdCIsImIhdCI6MTc1NzY0MjI0Mn0.PtuouxbW2bLVBCDaZBarPDovXu1TqCYJ2-t9sdSxL4M>



Gracias

Gracias!

Email

email@tec.mx