ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA EN SISTEMAS
RECUPERACIÓN DE INFORMACIÓN | 2025-B

Proyecto de 2do Bimestre
Sistema de Recuperación de Información

Nombres: Riki Guallichico; Kevin Martínez

1. Descripción del Corpus y Preprocesamiento

Para la construcción del sistema, se seleccionó el dataset "Consumer Reviews of Amazon Products" disponible en Kaggle, cumpliendo con los requisitos de indexación multimodal. Con el objetivo de robustecer el catálogo, se implementó una estrategia de fusión de datos que concatena dos fuentes primarias del dataset (*Datafiniti_Amazon_Consumer_Reviews_of_Amazon_Products_May19.csv* y su versión complementaria). Esta unificación permitió consolidar un volumen mayor de productos y opiniones de usuarios para el entrenamiento del sistema.

El procesamiento de estos datos se llevó a cabo mediante un pipeline ETL (Extract, Transform, Load) desarrollado en Python. En esta etapa, el sistema normaliza los identificadores de producto (ASIN) y ejecuta un filtrado estricto de las URLs de las imágenes, descartando enlaces rotos o fotografías de baja resolución para asegurar la calidad visual de la interfaz. Un aspecto crucial de este preprocesamiento fue la creación de un campo de "contexto enriquecido" (*rag_context*), el cual concatena el título, la marca y un resumen de las reseñas más relevantes. Esta estructura de datos fue diseñada específicamente para dotar al modelo generativo de opiniones reales y evidencia textual, facilitando la justificación de las recomendaciones en la etapa final.

2. Arquitectura del Pipeline de Recuperación

El sistema se fundamenta en una arquitectura secuencial que integra búsqueda vectorial, reordenamiento neuronal y generación de texto. El proceso inicia con la codificación multimodal utilizando el modelo CLIP (*clip-ViT-B-32*). La elección de CLIP responde a su capacidad para proyectar tanto texto como imágenes en un mismo espacio vectorial compartido, lo que habilita la funcionalidad de búsqueda cruzada donde una consulta de texto puede recuperar imágenes y viceversa. Estos vectores se almacenan en ChromaDB, permitiendo una recuperación inicial eficiente de los candidatos más relevantes mediante similitud de coseno.

Tras la recuperación inicial de los candidatos (Top-K), el sistema aplica una etapa crítica de refinamiento conocida como Re-ranking. A diferencia de la búsqueda inicial que prioriza la velocidad, esta fase utiliza un modelo Cross-Encoder (*ms-marco-MiniLM-L-6-v2*) para analizar en profundidad la relación semántica entre la consulta del usuario y la descripción completa del producto. El Cross-Encoder asigna un puntaje de relevancia más preciso que reordena la lista de

resultados, asegurando que los productos mostrados al usuario estén semánticamente alineados con su intención de búsqueda y no solo visualmente similares.

La fase final del pipeline implementa la Generación Aumentada por Recuperación (RAG) utilizando el modelo Google Gemini 2.5 Flash. El sistema construye un prompt complejo que inyecta la información de los productos reordenados y el historial de la conversación. Esto permite al modelo actuar como un asistente experto que no solo presenta los productos, sino que justifica su selección basándose en las especificaciones técnicas y las reseñas de los usuarios procesadas anteriormente, cumpliendo con el requisito de *grounded generation* o generación fundamentada.

3. Análisis de Casos y Resultados

La eficacia del sistema se evaluó mediante diversos escenarios de búsqueda. En consultas textuales, como la búsqueda de "pilas recargables de larga duración", se observó cómo el sistema recupera inicialmente una variedad de baterías. Sin embargo, es el mecanismo de re-ranking el que discrimina positivamente, elevando a las primeras posiciones aquellos productos cuya descripción textual confirma explícitamente la característica "recargable", desplazando a baterías desechables que visualmente eran similares. El modelo generativo complementa este resultado explicando al usuario que la recomendación se basa en reseñas que confirman la durabilidad de la carga.

En escenarios de búsqueda visual, donde el usuario sube una imagen de referencia (por ejemplo, un dispositivo *Amazon Echo*), el sistema aprovecha los embeddings visuales de CLIP para identificar productos con morfología y empaquetado similar. La respuesta generada por el módulo RAG demuestra capacidad para reconocer la categoría del producto a partir de la imagen y ofrecer detalles técnicos pertinentes, validando la integración multimodal. Además, el sistema mantiene la coherencia en consultas de refinamiento; si el usuario solicita un cambio de color o característica tras una búsqueda inicial, el asistente conserva el contexto del producto anterior y filtra los resultados sin perder el hilo de la conversación.

4. Análisis Cualitativo del Impacto Técnico

El análisis cualitativo revela que la incorporación del Re-ranking transforma significativamente la experiencia de usuario en comparación con una búsqueda vectorial simple. Mientras que la búsqueda basada solo en embeddings (*Bi-encoder*) es rápida, tiende a fallar en consultas ambiguas donde el parecido visual no implica relevancia semántica. El Cross-Encoder mitiga este problema actuando como un filtro de calidad, aunque introduce una ligera latencia en el proceso, el intercambio por una mayor precisión justifica su implementación.

Por otro lado, la calidad de las respuestas generadas por el módulo RAG destaca por su nivel de fundamentación. Al forzar al modelo a utilizar exclusivamente el contexto recuperado del dataset (*rag_context*), se reducen drásticamente las "alucinaciones" o información inventada. Las respuestas no son descripciones genéricas, sino que citan características específicas y opiniones de compradores reales presentes en el corpus, lo que otorga mayor credibilidad y utilidad al asistente conversacional.

5. Discusión Crítica, Limitaciones y Trabajo Futuro

A pesar de que el sistema cumple con el flujo de recuperación multimodal, se han identificado limitaciones técnicas y errores sistemáticos durante las pruebas. Una de las principales deficiencias observadas radica en el "colapso semántico" del modelo CLIP en situaciones de alta similitud visual. En ciertos casos, el sistema recupera productos que visualmente son casi idénticos a la consulta (por ejemplo, una funda de teléfono con estampado de calculadora) pero funcionalmente distintos al objeto buscado (una calculadora real). Aunque el mecanismo de re-ranking mitiga este error al analizar el texto, depende enteramente de que el paso inicial de recuperación (retrieval) haya logrado incluir al menos un candidato correcto en el conjunto preliminar; si el modelo visual falla en traer el producto correcto al Top-20, el re-ranker no tiene materia prima útil para corregir el resultado.

Otra limitación significativa es el compromiso entre precisión y latencia introducido por la arquitectura de dos etapas. La implementación del Cross-Encoder, si bien mejora drásticamente la relevancia de los resultados textuales, añade una sobrecarga computacional perceptible, incrementando el tiempo de respuesta en aproximadamente 300 milisegundos por consulta. En un entorno de producción con miles de usuarios concurrentes, esta arquitectura requeriría optimización mediante destilación de modelos o el uso de hardware dedicado (GPU) para mantener la experiencia de usuario fluida, ya que la ejecución actual en CPU puede volverse un cuello de botella.

Finalmente, se identifican áreas claras para mejoras futuras que robustecerían el sistema. La incorporación de una "búsqueda híbrida" que combine la similitud vectorial con algoritmos de palabras clave tradicionales (como BM25) ayudaría a resolver búsquedas de términos exactos (como números de modelo específicos) donde los embeddings a veces pierden precisión. Asimismo, el módulo RAG podría beneficiarse de una ventana de contexto más amplia que permita comparar más de cinco productos simultáneamente, ofreciendo al usuario una tabla comparativa generada en tiempo real en lugar de una recomendación lineal. Estas mejoras transformarían el prototipo actual en una herramienta comercialmente viable.

6. Evidencia:

Configuración

Imagen de referencia

Drag and drop file here

Limit 200MB per file • JPG, PNG, JPEG

Browse files

Limpiar Sesión

Stop Deploy

Asistente de Compras Inteligente

Recomiendame jaulas

Buscando: **'cages pets'** (Texto Inteligente)...

Debug: Score Top: 0.2726 | Umbral: 0.15

¡5 encontrados!

Pensando...

¿Qué buscas hoy?

Configuración

Imagen de referencia

Drag and drop file here

Limit 200MB per file • JPG, PNG, J...

Browse files

Limpiar Sesión

Deploy

Recomiendame jaulas

Buscando: **'cages pets'** (Texto Inteligente)...

Debug: Score Top: 0.2726 | Umbral: 0.15

¡5 encontrados!

Le recomiendo la **AmazonBasics Double-Door Folding Metal Dog Crate - Medium (36x23x25 Inches)**. Es una jaula resistente y bien valorada por los clientes, con una calificación de 4/5 estrellas. Cuenta con dos puertas para mayor comodidad.

AmazonBasics Double-Door Folding Metal Dog Crate - Medium (36x23x25 Inches)

¿Qué buscas hoy?

Configuración

Imagen de referencia

Drag and drop file here

Limit 200MB per file • JPG, PNG, J...

Browse files

Limpiar Sesión

muestrame las que mejor calificaciones tiene

Buscando: **'product pets'** (Texto Inteligente)...

Debug: Score Top: 0.2158 | Umbral: 0.15

¡5 encontrados!

Claro. Le recomiendo la **Two Door Top Load Pet Kennel Travel Crate Dog Cat Pet Cage Carrier Box Tray 23'**, con una calificación perfecta de 5/5 estrellas. Los clientes destacan su robustez, ligereza y facilidad de limpieza.

Cat Litter Box Covered Tray Kitten Extra Large Enclosed Hooded Hidden Toilet

Rel: 0.22

¿Qué buscas hoy?