



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
CIENCIAS DE LA COMPUTACIÓN

---

# Proyecto Recuperación de Información 2do Bimestre

Alexis Vera

Sistema de Recuperación  
Multimodal de Información



## 1. Introducción

Este informe presenta el desarrollo de un proyecto de Recuperación de Información, sobre un Sistema de Recuperación Multimodal de Información en un escenario de e-commerce, enfocado en la búsqueda de productos de una tienda online tipo Amazon mediante una entrada de texto o una imagen, su objetivo principal es permitir encontrar productos similares a la búsqueda inicial utilizando representación semántica.

El proyecto prioriza, la estabilidad del entorno y la correcta integración de modelos preentrenados con mecanismos de indexación vectorial.

## 2. Objetivo

Diseñar e implementar un sistema de Recuperación de Información Multimodal aplicado a un escenario de e-commerce, que integre búsqueda multimodal, re-ranking y generación aumentada por recuperación (RAG).

- Implementar un modelo de embeddings semánticos utilizando CLIP.
- Construir un índice vectorial eficiente con FAISS.
- Permitir la búsqueda tanto textual como visual.
- Integrar el sistema en una aplicación web interactiva con Streamlit.

## 3. Descripción del Sistema

Empleando una arquitectura de recuperación semántica, los textos y las imágenes son transformados en vectores numéricos dentro de un mismo espacio de representación.

Estos vectores son almacenados en un índice FAISS, lo que permite realizar búsquedas rápidas por similitud utilizando distancia vectorial y que ambos flujos al utilizar el mismo índice vectorial, garantiza la coherencia en los resultados.



## 4. Descripción del corpus

El corpus utilizado fue adquirido desde la plataforma Kaggle (Consumer Reviews of Amazon Products) contando con cerca de 5000 registros que representan un producto e incluye información textual y enlaces para representaciones visuales relevantes para la recuperación.

Luego de un preprocesamiento y limpieza del corpus nos quedamos con las siguientes columnas:

- **id**: Identificador único del producto.
- **name**: Nombre del producto.
- **categories / primaryCategories**: Categorías asociadas al producto.
- **reviews.text**: Opiniones y reseñas de usuarios.
- **text / product\_text / combined\_text**: Campos textuales utilizados para la representación semántica.
- **image\_url**: Enlace a la imagen representativa del producto.

Este corpus posibilita el trabajo multimodal porque integra información visual y textual, ambas esenciales para la recuperación.

## 5. Pipeline del sistema

El pipeline del proyecto se divide en tres etapas claves:

### 5.1. Retrieval (Recuperación)

- Durante la etapa de recuperación, se convierten en embeddings vectoriales las imágenes y los textos, empleando el modelo CLIP de OpenAI. Estos embeddings son normalizados y se comparan con un índice vectorial que fue creado con FAISS.
- FAISS posibilita hacer búsquedas eficaces por similitud, devolviendo en el espacio semántico los k artículos más próximos.



## 5.2. Re-ranking

- Después de la recuperación inicial, se pueden reorganizar los resultados (re-ranking) con base en criterios adicionales como:
  - Más semejanza en términos de significado.
  - Coincidencia con las categorías primordiales.
  - Pertinencia contextual del texto combinado.

El re-ranking prioriza los productos que tienen una mayor coherencia con la intención de la consulta, lo cual mejora la calidad de los resultados finales.

## 5.3. RAG

- Los productos que se han recuperado son utilizados por el sistema RAG como contexto para mejorar la respuesta final al usuario. Esta estructura arquitectónica posibilita expandir el sistema con el fin de producir respuestas explicativas fundamentadas en los productos hallados.

# 5. Ejemplos de consultas y resultados obtenidos

- Ejemplo 1:

```
▶ query = "smart speaker with voice assistant"

results = text_search_with_reranking(query)
context = build_rag_context(results)

prompt = build_rag_prompt(query, context)
recommendation = generate_recommendation(prompt)

print(recommendation)

*** Basado en tu consulta de "smart speaker with voice assistant" y la
      * **Amazon Tap - Alexa-Enabled Portable Bluetooth Speaker**

Este producto es un "speaker" y es "Alexa-Enabled", lo que indica q
```



- Ejemplo 2:

```
▶ query = "amazon echo speaker"

results = text_search_with_reranking(query)
context = build_rag_context(results)

prompt = build_rag_prompt(query, context)
recommendation = generate_recommendation(prompt)

print(recommendation)

*** Te recomiendo el producto:

* **Amazon - Echo Plus w/ Built-In Hub -
```

- Ejemplo 3:

```
▶ query = "amazon bluetooth device"

results = text_search_with_reranking(query)
context = build_rag_context(results)

prompt = build_rag_prompt(query, context)
recommendation = generate_recommendation(prompt)

print(recommendation)

*** Basado en tu consulta "amazon bluetooth device" y la información proporcionada, te
    * **Amazon Tap - Alexa-Enabled Portable Bluetooth Speaker**

    **Justificación:**  

Este producto es adecuado porque su nombre incluye "Amazon" y "Bluetooth Speaker",
```

## 5. Análisis cualitativo y calidad de las respuestas

La calidad de los resultados finales se ve afectada significativamente por el re-ranking. Los resultados que se obtienen pueden ser semánticamente similares, pero menos relevantes para el usuario, sin esta etapa

- Se mejora la precisión en los primeros resultados.
- Se reduce la aparición de productos ambiguos.
- Se incrementa la coherencia entre consulta, texto e imagen.

El proceso de recuperación tiene un impacto directo en la calidad de las respuestas producidas por el sistema RAG. El sistema ofrece contextos pertinentes para la generación al emplear embeddings multimodales y un índice vectorial eficaz.

## 6. Conclusiones

- El sistema desarrollado muestra la viabilidad de una perspectiva multimodal RAG que se basa en CLIP y FAISS, cumpliendo con los objetivos planteados. Integrando apropiadamente el pipeline de recuperación, re-ranking y generación.
- Este proyecto establece una base firme para futuras ampliaciones enfocadas en generar un lenguaje natural más avanzado y en mejorar el proceso de recuperación en situaciones de mayor envergadura.

## 7. Recomendaciones

- Se aconseja ampliar el conjunto de datos incluyendo una mayor gama de productos para identificar una mayor cantidad de textos e imágenes.
- Implementar una búsqueda combinada de texto e imagen.

## 8. Bibliografía

Dataset. (2019). Consumer Reviews of Amazon Products. Kaggle.

<https://www.kaggle.com/datasets/datafiniti/consumer-reviews-of-amazon-products/data>