|  |
| --- |
| Project Report  Title: Myntra Product Search App: Semantic Search and Question Answering for Fashion Products  Prepared By: Anand Vardhan Sharan  Date: 28-Apr-2024 |

Table of Contents

[Overview 3](#_Toc165164241)

[1. Introduction 3](#_Toc165164242)

[2. Scope and Requirements 4](#_Toc165164243)

[3. System Diagram 5](#_Toc165164244)

[4. Challenges & Lessons Learned 5](#_Toc165164245)

[5. Sample Screenshots 6](#_Toc165164246)

[5. Conclusion 7](#_Toc165164247)

# Overview

Problem Statement: Customers often struggle to find specific fashion products they are looking for on e-commerce sites like Myntra due to limitations of keyword search.

Solution: The Myntra Product Search App combines semantic search with AI to provide accurate and contextual results for user queries on the Myntra fashion product dataset.

Core Features:

* Data preprocessing from CSV files
* Semantic search to retrieve relevant product information
* AI-powered question answering to directly show relevant product details

Benefits: Enhances product discovery, saves time for customers, and potentially increases sales by showing more relevant products.

# 1. Introduction

Problem Statement:

Keyword search is limited in its ability to truly understand the user's intent behind a search query, especially for domains like fashion that are rich in descriptive terminology. There are several key issues that arise:

1. Inability to Handle Synonyms and Paraphrasing

Fashion product descriptions use a wide variety of synonymous terms to describe attributes like styles, fits, materials, etc. For example, "slim fit", "skinny", and "tight" could all be used to describe a fitted clothing item. Keyword search cannot inherently make these connections between synonymous phrases.

2. Lack of Context Understanding

A keyword like "red" could refer to the color of a dress, or it could indicate the user wants a dress for a red carpet event. Keyword search cannot decipher the intended context and meaning behind individual words in a query.

3. Challenges with Multiword Phrases

Product attributes are often described using multiword phrases like "balloon sleeve top" or "high-waisted jeans". Keyword search may miss relevant results if it fails to recognize the entire phrase as a single concept.

4. No Query Expansion

If a user searches for "summer dresses", keyword search cannot automatically expand and incorporate other related terms like "sundresses", "beach dresses", "warm weather dresses" etc. into the search process.

5. No Handling of Misspellings or Typographic Errors

Keyword search strictly matches the input keywords, so any misspelled words or typographic errors in the user's query will fail to retrieve intended results.

In contrast, semantic search combined with natural language processing can help overcome these limitations by understanding the contextual meaning, handling synonyms, breaking down queries into semantic concepts, and providing more relevant results aligned with the user's true intent behind the search query.

Objectives:

* Preprocess the Myntra product dataset into an optimized format
* Implement semantic search for accurate product retrieval
* Develop natural language query capabilities with question answering

Target Audience: Customers shopping for fashion products on Myntra or similar e-commerce platforms.

# 2. Scope and Requirements

Features:

* Data Preprocessing: CSV parsing, text cleaning, column combining
* Semantic Search: LangChain, OpenAI embeddings, vector stores
* Question Answering: Integration with OpenAI language models
* Web UI: Gradio-based user interface

Technology Stack:

- Python

- pandas, NumPy: Data preprocessing

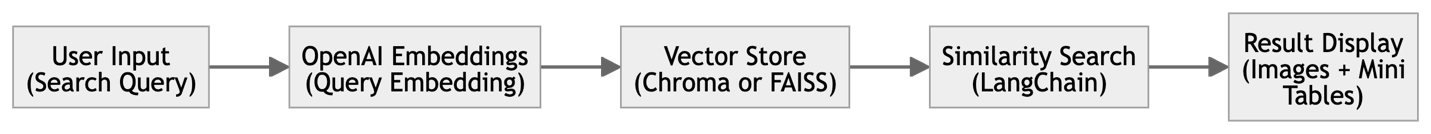
- OpenAI Text embeddings, language models

- LangChain: Framework for semantic search & QA

- Gradio: Web UI creation

- Other Libraries: ast, pathlib, BeautifulSoup, PIL

# 3. System Diagram



# 4. Challenges & Lessons Learned

Challenges:

* Product Data Structure: Handling varied product attributes
* Query Understanding: Mapping natural language to product specs
* Result Ranking: Providing most relevant products first
* Performance: Fast retrieval from large dataset

Lessons Learned:

- Data Preprocessing: Critical for usable vector embeddings

- Query Adaptation: Adjusting models to e-commerce domain

- Embedding Storage: Vector stores enable fast similarity search

- UI/UX Design: Intuitive display of text and visuals

# 5. Sample Screenshots

Example 1:

A screenshot of a computer

Description automatically generated

Example 2:

A screenshot of a computer

Description automatically generated

# 5. Conclusion

Summary:

* The Myntra Product Search App showcases the power of combining semantic search and natural language processing to enhance e-commerce product discovery.
* Key innovations include data preprocessing, embedding-based retrieval, and intuitive UI.

Future Work:

* Deep product attribute filtering
* Multimodal search with images
* Recommendations based on user history.