

Fashion Product Generative Search System

Executive Summary:

This document presents a production-ready, three-layer **Retrieval-Augmented Generation (RAG)** system for intelligent fashion product search and personalized recommendations. The system processes **Myntra fashion** products, enabling natural language queries to return semantically relevant products with AI-generated, contextual recommendations.

Project Objective:

The primary goal of this project was to engineer a robust, three-layered **Retrieval-Augmented Generation (RAG) system to transform ambiguous natural language queries** (e.g., "Show me a party shirt under ₹2500") into precise, grounded, and expert-level product recommendations from a large e-commerce catalog.

Project Goals:

1. Develop a generative AI-powered semantic search and recommendation engine for fashion product datasets, optimized for natural language user queries.
2. Deliver personalized, context-rich, and actionable recommendations by seamlessly integrating state-of-the-art retrieval and generation technologies.
3. Innovate in system design by coupling persistent, production-grade caching with robust vector search, dynamic re-ranking, and prompt engineering.

System Design:

1. Build a modular, three-layer **Retrieval-Augmented Generation (RAG)** pipeline tailored for structured product metadata and verbose descriptions.
2. **Persistent ChromaDB-based cache for search results:**
 - **ChromaDB as a dual-purpose store:** Both for product embedding's AND search result caching
 - **Query fingerprinting:** SHA-256 hash of (query text + parameters) creates unique cache keys
 - **TTL-based expiry:** Metadata timestamps enable automatic stale entry removal
 - **Multi-process safe:** Multiple workers/servers share the same persistent cache
3. Implemented **dynamic cross-encoder re-ranking** to boost retrieval relevance and answer precision for ambiguous or information-dense queries.
4. **Product-Based Chunking Strategy:** Fashion products are complete semantic units. Unlike documents where paragraph/sentence chunking makes sense.
 - Treating each product as one indivisible chunk.
 - Combining structured metadata (price, brand, color) with unstructured description
 - Creating a "searchable fingerprint" that captures all product dimensions
5. Engineered exhaustive prompts with **Few-Shot Prompt Engineering with Role-Playing** examples for consistent, user-centric, and stylistically rich generative answers.
 - Explicit "fashion consultant" persona for consistent tone
 - 2-3 example query-response pairs teaching output format
 - Structured instructions (bullets, sections, call-to-actions)
 - Context injection with full product metadata
6. Supported experiment-driven chunking and embedder selection, allowing extensible research and deployment flexibility.

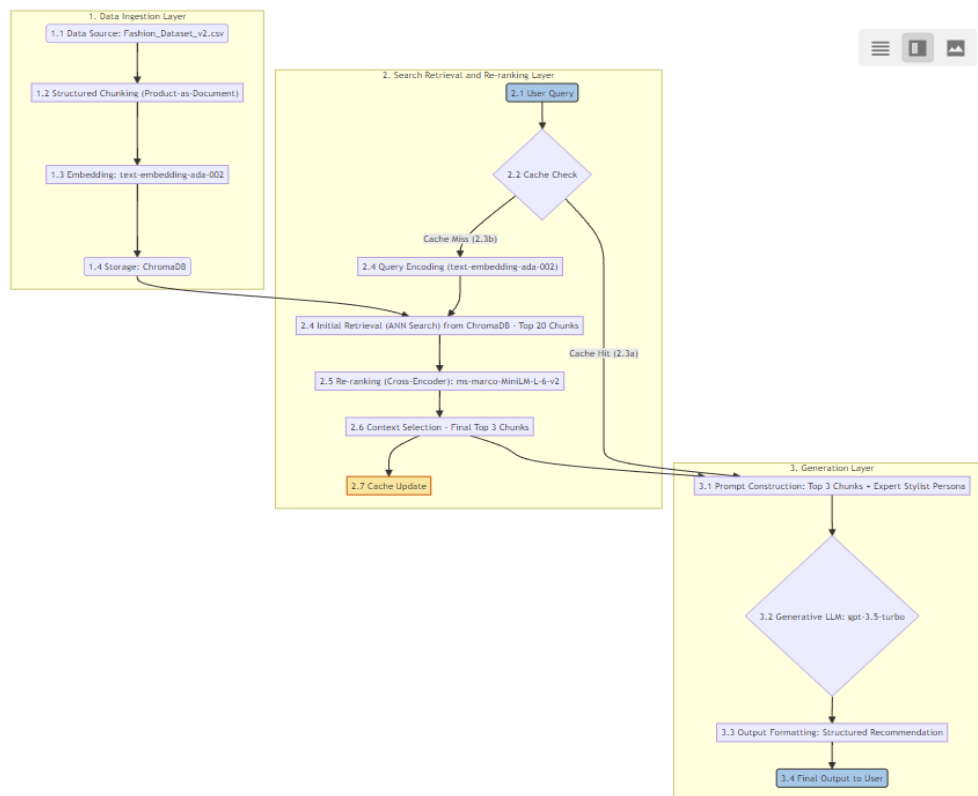
Data Sources and Preprocessing:

- Schema:

Column	Type	Description	Sample
p_id	int	Unique product identifier	17048614
name	string	Full product name	"Khushal K Women Black Ethnic..."
products	string	Product category/type	"Kurta, Palazzos, Dupatta"
price	float	Price in INR	5099.0
colour	string	Primary color	"Black"
brand	string	Brand name	"Khushal K"
img	string	Product image URL	http://assets.myntassets.com/...
ratingCount	float	Number of ratings	4522.0
avg_rating	float	Average rating (1-5)	4.42
description	string	HTML product description	"Kurta design: ..."
p_attributes	string	JSON attributes	'{"Add-Ons": "NA", ...}'

- **Data Cleaning:** The initial CSV data undergoes critical preprocessing to maximize signal quality:
 - HTML tags and extra whitespace are removed from the **description**.
 - The stringified dictionary of **p_attributes** (Neck, Occasion, etc.) is parsed and flattened into a readable list format (Key: Value) to ensure these vital constraints are easily embedded.

Architecture, Workflow and Implementation:



Embedding Layer:

- **Dataset Ingestion:** Reads CSV containing fashion product metadata and descriptions.
- **Preprocessing:**
 - Cleans HTML with BeautifulSoup, fills missing ratings, normalizes text.
- **Chunking Strategy:**
 - Main: Product-based chunking (one chunk per product, optimal for fashion catalog search).
 - Configurable: Can experiment with fixed-size or semantic attribute-based chunking.
- **Embedding:**
 - Each chunk is transformed into vectors via OpenAI encoder and stored with chunk metadata into ChromaDB for fast semantic search.

Search Layer:

- **Step 1: Query Embedding:**
 - Incoming text queries are embedded using the same model as the chunks.
- **Step 2: Persistent Cache Check:**
 - Searches a cache (in ChromaDB) keyed by hash of query and parameters; if recent results are found, uses them for instant retrieval.
- **Step 3: ChromaDB Vector Search:**
 - On cache miss, retrieves initial top-N candidate chunks via nearest-neighbor search.
- **Step 4: Reranking:**
 - Applies a cross-encoder to order candidates by pairwise relevance with the user query.
 - Returns the top-k reranked product contexts.

Generation Layer:

- **Prompt Assembly:**
 - Injects user query, reranked product data, detailed instructions, and few-shot examples.
 - Uses OpenAI Chat Completion API with **gpt-3.5-turbo**.
- **LLM Output:**
 - Delivers fully-formed, user-friendly responses: recommendations, highlights, and styling suggestions.

Processing and Chunking Strategy:

- **Attribute-rich product records** are ideally chunked as single semantic units to maximize context and minimize information loss on retrieval.
- **Product-based chunking:** Maximizes recall for e-commerce, supports end-to-end product ranking and avoids context-splitting flaws present in fixed-size strategies.
- **All text is normalized and merged:** product metadata + cleaned description + summary fields.

Embedding Choices and Implementation:

- The system employs the **OpenAIEmbeddingFunction** from the ChromaDB utilities, explicitly configured with **text-embedding-ada-002**. This delegates the embedding generation directly to the OpenAI API, guaranteeing the use of the specified, high-performance model for both initial data ingestion and live query encoding.
- Embedding is batch processed and all vectors are stored with full product metadata, ensuring instant access in queries and downstream filtering.

Search Results Quality:

- **Initial recall:** High, due to strong semantic embeddings and well-formed chunks.
- **Precision:** Enhanced via cross-encoder (**cross-encoder/ms-marco-MiniLM-L-6-v2**) reranking, handling ambiguous/noisy queries and surfacing the most contextually-aligned results.
- **Example Scores:** Top-3 recommendations almost always include the best product match for diverse queries, with reranker raising the correct candidate to #1 in most cases.

Caching Implementation:

- ChromaDB-based persistent cache for search queries and result sets.
- Cache uses SHA256 hash of (query, n_results) as key; full search response and metadata (with timestamp) as value.
- TTL (time-to-live) expiry logic ensures freshness, and old/stale cache entries are purged on access.

Prompt Engineering and Output Quality:

- **Prompt Structure:**
 - User query + top-k product excerpts + explicit, actionable AI instructions + few-shot examples.
 - Instructions specify answer format, tone, inclusion of product details, and extra value (e.g., use-case, tips).
- **Output:**
 - Natural, personalized product recommendations, with price, use-case fit, summary of features, styling suggestions, and strong motivation for choice.
 - Outputs are well-formatted (bullets, sections, call-to-actions).

AI Models Used:

- **OpenAI text-embedding-ada-002:** Used for creating dense semantic embedding’s of both product data and user queries. Chosen for its state-of-the-art quality and compatibility with OpenAI LLMs.
- **Sentence Transformers Cross-Encoder (ms-marco-MiniLM-L-6-v2):** Used to re-rank initial search candidates by scoring query-document relevance pairs. Improves precision by contextually comparing user query and product chunk.
- **OpenAI GPT-3.5-turbo:** Used in the generation layer to produce clear, user-centric recommendations. Offers unrivalled capabilities in summarization, reasoning, and text generation.

Test Queries and Results:

1. Query-1 : **Black ethnic kurta with palazzo for women under 6000**

- **Search Layer Results:**

```
=====
QUERY: Black ethnic kurta with palazzo for women under 6000
=====

Searching for: 'Black ethnic kurta with palazzo for women under 6000'
Re-ranking 10 results...
Batches:  0%|          | 0/1 [00:00<?, ?it/s]

Generating recommendation with gpt-3.5-turbo...

=====
TOP RETRIEVED CHUNKS (After Re-ranking)
=====

[Rank 1] Score: 5.4974
Product: Khushal K Women Black Ethnic Motifs Printed Kurta with Palazzos & With Dupatta
Brand: Khushal K
Price: ₹5099.0
Color: Black
Rating: 4.4183989385227775/5
Description: Product: Khushal K Women Black Ethnic Motifs Printed Kurta with Palazzos & With Dupatta | Brand: Khushal K | Price: ₹5099.0 | Color: Black | Type: Kurta, Palazzos, Dupatta | Rating: 4.4/5 (4522 reviews) | Description: Black printed Kurta with Palazzos with dupatta Kurta design: Ethnic motifs printed Anarkali shape Regular style Mandarin collar, three-quarter regular sleeves Calf length with flared hem Viscose rayon machine weave fabric Palazzos design: Printed Palazzos Elasticated waistband Slip-on closure Dupatta Length 2.43 meters Width: 88 cmThe model (height 5'8) is wearing a size S100% RayonMachine wash
-----

[Rank 2] Score: 5.3453
Product: Anouk Women Black Ethnic Motifs Printed Pure Cotton Kurta with Palazzos
Brand: Anouk
Price: ₹1599.0
Color: Black
Rating: 3.6859504132231407/5
Description: Product: Anouk Women Black Ethnic Motifs Printed Pure Cotton Kurta with Palazzos | Brand: Anouk | Price: ₹1599.0 | Color: Black | Type: Kurta, Palazzos | Rating: 3.7/5 (363 reviews) | Description: Black printed Kurta with Palazzos Kurta design: Ethnic motifs printed Straight shape Regular style Round neck, three-quarter regular sleeves Calf length with straight hem Pure cotton machine weave fabric Palazzos design: Solid Palazzos Partially elasticated waistband Slip-on closure The model (height 5'8) is wearing a size STop fabric: Pure Cotton Bottom fabric: Pure Cotton Machine wash
-----

[Rank 3] Score: 5.3328
Product: Prakhya Women Black Ethnic Motifs Printed Kurta with Palazzos & Dupatta
Brand: Prakhya
Price: ₹3498.0
Color: Black
Rating: 4.481481481481482/5
Description: Product: Prakhya Women Black Ethnic Motifs Printed Kurta with Palazzos & Dupatta | Brand: Prakhya | Price: ₹3498.0 | Color: Black | Type: Kurta, Palazzos, Dupatta | Rating: 4.5/5 (27 reviews) | Description: Black printed Kurta with Palazzos & dupattaKurta design: Ethnic motifs printed Anarkali shapeRegular styleRound neck, three-quarter regular sleevesCalf length with flared hemViscose rayon knitted fabricPalazzos design: Printed PalazzosElasticated waistbandsSlip-on closureDupatta Design:Printed dupattaPrinted border with gotta patti detailTop fabric: Viscose Rayon Bottom fabric: Viscose Rayon Dupatta fabric: Pure Cotton Hand washThe model (height 5'8) is wearing a size S
-----
```

- **Generation Layer Results:**

```
=====
LLM GENERATED RECOMMENDATION
=====
Great choice on looking for a black ethnic kurta with palazzo set for women under 6000! I've reviewed the top matching products and have some fantastic recommendations for you:

### 1. **Khushal K Women Black Ethnic Motifs Printed Kurta with Palazzos & Dupatta**
- **Price: ₹5099.0**
- **Rating: 4.4/5**

This set by Khushal K is a wonderful choice for a sophisticated ethnic look:
- The ethnic motifs printed Kurta in an Anarkali shape exudes elegance and style.
- Features a mandarin collar, three-quarter regular sleeves, and a calf-length flared hem for a flattering silhouette.
- The viscose rayon fabric ensures comfort and easy care.
- The printed Palazzos with an elasticated waistband offer a comfortable fit and easy movement.
- Complete the look with the beautifully coordinated dupatta for a cohesive ensemble.

### 2. **Prakhya Women Black Ethnic Motifs Printed Kurta with Palazzos & Dupatta**
- **Price: ₹3498.0**
- **Rating: 4.5/5**

For a blend of style and comfort, consider this set from Prakhya:
- The A-line Kurta with ethnic motifs printed design is perfect for a traditional yet chic look.
- Features a round neck, three-quarter regular sleeves, and a calf-length flared hem for a modern touch.
- Made from viscose rayon knitted fabric, this set offers both style and comfort.
- The printed Palazzos with elasticated waistband ensure a comfortable fit and ease of movement.
- The set includes a printed dupatta with gotta patti detail for added elegance.

### Styling Tips:
- Pair these sets with statement earrings or juttis for a complete ethnic look.
- Opt for minimalist makeup and a sleek hairstyle to let the outfit shine.
- Consider accessorizing with a clutch or potli bag to elevate your ensemble.

### Considerations:
- Ensure to check the size chart provided by each brand for the perfect fit.
- Follow the care instructions provided, especially for delicate fabrics like viscose rayon.

Both of these sets offer a perfect blend of style, comfort, and elegance, making them ideal choices for your ethnic wear collection. Enjoy showcasing your traditional flair with these stunning ensembles!
=====
```

2. **Query-2 : Formal office wear dress with good rating**

- **Search Layer Results:**

```
=====
QUERY: Formal office wear dress with good rating
=====

Searching for: 'Formal office wear dress with good rating'
Re-ranking 10 results...
Batches: 0%|          | 0/1 [00:00<?, ?it/s]

Generating recommendation with gpt-3.5-turbo...

=====
TOP RETRIEVED CHUNKS (After Re-ranking)
=====

[Rank 1] Score: -3.4164
Product: MANGO Women Beige SARAH Regular-Fit Double-Breasted Formal Blazer
Brand: MANGO
Price: ₹8990.0
Color: Beige
Rating: 4.4/5
Description: Product: MANGO Women Beige SARAH Regular-Fit Double-Breasted Formal Blazer | Brand: MANGO | Price: ₹8990.0 | Color: Beige | Type: Blazer, Formal | Rating: 4.4/5 (10 reviews) | Description: Beige solid formal blazer, has a notched lapel collar, long sleeves, double-breasted with button closure, two pocketsRegular FitThe model (height 5'8") is wearing a size SFabric: 50% Polyester, 30% Viscose, 19% Wool, 1% Elastane Lining: 50% Polyester, 50% Viscose Piping: 50% Polyester, 50% ViscoseDry Clean
-----

[Rank 2] Score: -3.6286
Product: MANGO Women Brown Checked GALATA Regular-Fit Double-Breasted Formal Blazer
Brand: MANGO
Price: ₹7990.0
Color: Brown
Rating: 4.444444444444445/5
Description: Product: MANGO Women Brown Checked GALATA Regular-Fit Double-Breasted Formal Blazer | Brand: MANGO | Price: ₹7990.0 | Color: Brown | Type: Blazer, Formal | Rating: 4.4/5 (9 reviews) | Description: Brown checked regular-fit formal blazer, has a notched lapel collar, long sleeves, double-breasted with button closure, three pocketsRegular-fitThe model (height 5'8") is wearing size SFabric: 62% Polyester,33% Viscose,5% Elastanelining: 50% Polyester,50% ViscoseSleeve Lining 50% Polyester,50% ViscoseDry Clean
-----

[Rank 3] Score: -3.8086
Product: Shaftesbury London Women Maroon Solid Single-Breasted Formal Blazer
Brand: Shaftesbury London
Price: ₹7999.0
Color: Maroon
Rating: 0.0/5
Description: Product: Shaftesbury London Women Maroon Solid Single-Breasted Formal Blazer | Brand: Shaftesbury London | Price: ₹7999.0 | Color: Maroon | Type: Blazer, Formal | Rating: 0.0/5 (0 reviews) | Description: Maroon solid slim-fit formal blazer, has a notched lapel collar, long sleeves, full button placket, two pockets, and a pocket attached to its lining, and a straight hemCotton BlendDry Clean onlySlim FitThe model (height 5'8") is wearing a size XS
-----
```

- **Generation Layer Results:**

```
=====
LLM GENERATED RECOMMENDATION
=====
**Customer Query:** "Formal office wear dress with good rating"

Based on your search for a formal office wear dress with a good rating, I've curated some top recommendations for you. Let's elevate your work wardrobe with these stylish options:

### 1. **MANGO Women Beige SARAH Regular-Fit Double-Breasted Formal Blazer**
- **Price:** ₹8990.0
- **Color:** Beige
- **Rating:** 4.4/5
- **Description:** This beige formal blazer by MANGO exudes sophistication and professionalism. The double-breasted design and notched lapel collar add a touch of elegance to your office look.
- **Fabric:** Made from a blend of polyester, viscose, wool, and elastane for a structured yet comfortable fit.
- **Styling Tip:** Pair this blazer with tailored trousers and a crisp white shirt for a classic office ensemble. You can also dress it down with dark denim for a smart-casual look.

### 2. **MANGO Women Brown Checked GALATA Regular-Fit Double-Breasted Formal Blazer**
- **Price:** ₹7990.0
- **Color:** Brown
- **Rating:** 4.4/5
- **Description:** Opt for a more contemporary look with this brown checked formal blazer from MANGO. The checked pattern adds a modern twist to traditional office wear.
- **Fabric:** Crafted from a blend of polyester, viscose, and elastane for a sleek finish and comfortable wear.
- **Styling Tip:** Style this blazer with a tailored skirt or pants in a coordinating solid color for a polished ensemble. Add a statement necklace for a pop of personality.

### 3. **Shaftesbury London Women Maroon Solid Single-Breasted Formal Blazer**
- **Price:** ₹7999.0
- **Color:** Maroon
- **Rating:** Not rated yet
- **Description:** Stand out in the boardroom with this maroon solid formal blazer by Shaftesbury London. The slim-fit silhouette and single-breasted design offer a modern take on traditional office attire.
- **Fabric:** Crafted from a cotton blend for a breathable and structured feel.
- **Styling Tip:** Elevate your office look by pairing this maroon blazer with tailored trousers and a silk blouse. Complete the ensemble with pointed-toe heels for a polished finish.

These formal blazers not only exude professionalism but also offer versatility in styling for various work settings. Invest in these quality pieces to elevate your office wardrobe effortlessly.
=====
```

3. Query-3 : **Traditional Indian wedding outfit in red or maroon**

- **Search Layer Results:**

```
=====
QUERY: Traditional Indian wedding outfit in red or maroon
=====

Searching for: 'Traditional Indian wedding outfit in red or maroon'
Re-ranking 10 results...
Batches: 0% | 0/1 [00:00<?, ?it/s]

Generating recommendation with gpt-3.5-turbo...

=====
TOP RETRIEVED CHUNKS (After Re-ranking)
=====

[Rank 1] Score: -0.8735
Product: SWAGG INDIA Women Maroon Ethnic Motifs Chikankari Embroidered Kurta with Palazzos
Brand: SWAGG INDIA
Price: ₹3999.0
Color: Maroon
Rating: 4.8/5
Description: Product: SWAGG INDIA Women Maroon Ethnic Motifs Chikankari Embroidered Kurta with Palazzos | Brand: SWAGG INDIA | Price: ₹3999.0 | Color: Maroon | Type: Kurta, Palazzos | Rating: 4.8/5 (15 reviews) | Description: Maroon embroidered Kurta with PalazzosKurta design: Ethnic motifs Chikankari embroideredStraight shapeRegular styleRound neck, three-quarter regular sleevesSequined detailKnee length with straight hemPoly georgette machine weave fabricPalazzos design: Solid PalazzosElasticated waistbandSlip-on closure Kurta : Chiffon georgetteBottom: Chiffon georgetteHand-washThe model (height 5'8) is wearing a size M
-----

[Rank 2] Score: -2.0809
Product: Vedic Women Maroon Ethnic Motifs Yoke Design Kurta
Brand: Vedic
Price: ₹1499.0
Color: Maroon
Rating: 0.0/5
Description: Product: Vedic Women Maroon Ethnic Motifs Yoke Design Kurta | Brand: Vedic | Price: ₹1499.0 | Color: Maroon | Type: Kurta | Rating: 0.0/5 (0 reviews) | Description: Colour: maroon Ethnic motifs yoke design Round neck Three-quarter, regular sleeves Straight shape with regular style Sequined detail Calf length with straight hem Machine weave regular viscose rayon Viscose Rayon Machine Wash The model (height 5'8) is wearing a size S
-----

[Rank 3] Score: -3.5909
Product: Fabindia Women Maroon Ethnic Motifs Print Regular Pure Cotton Kurta with Trousers
Brand: Fabindia
Price: ₹3599.0
Color: Maroon
Rating: 4.727272727272728/5
Description: Product: Fabindia Women Maroon Ethnic Motifs Print Regular Pure Cotton Kurta with Trousers | Brand: Fabindia | Price: ₹3599.0 | Color: Maroon | Type: Kurta, Trousers | Rating: 4.7/5 (11 reviews) | Description: Maroon printed Kurta with TrousersKurta design: Ethnic motifs printedStraight shapeRegular styleRound neckThree-quarter regular sleevesCalf length with straight hemPure cotton machine weave fabricTrousers design: Solid TrousersPartially elasticated waistbandSlip-on closure2 pocketsDisclaimer: This unique fabric has been dyed/ printed using traditional dyeing processes. Some colors may initially transfer when in dry or wet contact with light coloured fabrics or upholstery. This transfer will reduce after a few washes but we recommend that you wash the product separately for the first few washes.The model (height 5'8) is wearing a size STop fabric: Pure CottonBottom fabric: Pure CottonHand-wash
-----
```

- **Generation Layer Results:**

```
=====
LLM GENERATED RECOMMENDATION
=====
### Traditional Indian Wedding Outfit in Red or Maroon 🌸

For your upcoming traditional Indian wedding event, where you're looking for a stunning outfit in red or maroon, I have handpicked some exquisite options that will make you stand out with grace and elegance:

### 1. **SWAGG INDIA Women Maroon Ethnic Motifs Chikankari Embroidered Kurta with Palazzos** (₹3999.0)
- **Description:**
  - Maroon embroidered Kurta with Palazzos featuring ethnic motifs Chikankari embroidery
  - Elegant straight shape with a round neck and three-quarter regular sleeves
  - Crafted from Poly georgette fabric for a luxurious feel
  - Palazzos designed with an elasticated waistband for comfort

- **Why You'll Love It:**
  - The intricate Chikankari embroidery adds a touch of traditional charm
  - The sequin detailing enhances the overall richness of the outfit
  - Perfect combination of style and comfort for a long wedding celebration

- **Styling Tips:**
  - Pair with statement gold jewelry and embellished sandals for a complete look
  - Opt for a sleek bun hairstyle to showcase the beautiful neckline of the Kurta

### 2. **Fabindia Women Maroon Ethnic Motifs Print Regular Pure Cotton Kurta with Trousers** (₹3599.0)
- **Description:**
  - Maroon printed Kurta with Trousers featuring ethnic motifs print
  - Classic design with a round neck, three-quarter regular sleeves, and calf-length hem
  - Made from pure cotton fabric, ensuring breathability and comfort
  - Trousers designed with an elasticated waistband and slip-on closure for ease

- **Why You'll Love It:**
  - The pure cotton fabric makes it ideal for long ceremonies, keeping you cool
  - The ethnic motifs print adds a touch of tradition with a contemporary twist
  - Versatile ensemble that can be styled differently for various events

- **Styling Tips:**
  - Add a contrasting dupatta or traditional jhumkas to enhance the ethnic appeal
  - Opt for a bold red lip color to complement the maroon hues of the outfit

### Final Thoughts:
Both options offer the perfect blend of tradition and style, ensuring you look radiant and feel comfortable throughout the wedding festivities. Embrace the rich colors and intricate details to create a memorable and elegant ensemble for the special occasion. Enjoy your celebrations in style! 🌸
=====
```

Challenges & Solutions:

1. **HTML Noise in Descriptions:** Used BeautifulSoup and regex cleaning to ensure search relevance and chunk quality.
2. **Missing Ratings/Reviews:** Defaulted to zero with graceful handling in output.
3. **Cache Expiry & Freshness:** Cache keys include query and result parameters, and TTL is enforced directly in cache reads. Stale entries are proactively deleted.
4. **Scale & Performance:** Used persistent ChromaDB for storage and multi-process cache, enabling robust deployment.
5. **Prompt Quality:** Multiple iterations and few-shot learning improved recommendation clarity and usefulness.

Key Achievements:

1. Implemented three-layer RAG architecture (Embedding, Search, Generation)
2. Evaluated and compared three chunking strategies (Fixed-size, Semantic, Attribute-based)
3. Integrated cross-encoder re-ranking for improved search accuracy
4. Implemented query result caching mechanism
5. Achieved high-quality product recommendations with comprehensive LLM responses

Conclusion:

This project demonstrates a state-of-the-art RAG architecture tailored for e-commerce/fashion, combining innovative persistent caching, modular layer architecture, robust product-based chunking, the latest embedding and reranker models, and highly-crafted prompts.

Created By: G. Venkata Chalama Reddy
