

INFORMATION RETRIEVAL AND EXTRACTION

FINAL REPORT

Abhay Kaushik
2022201054

Anshul Sharma
2023701011

Yash Sampat
2022201038

TEAM NO: 5

PROJECT: Conversational Outfit Retrieval

PROBLEM STATEMENT:

1. Give fashion recommendations to the user based on the prompt based on the fashion article and blogs, along with the relevant results by using the product catalogue.
2. Experiment with different methods like zero-shot. Explore memory buffers to implement a conversational setting.

DATASET COLLECTION:

1. We have collected datasets for 2 different tasks:
 - The first task is the relevant fashion blogs and articles for using in retrieval augmented generation.
 - The second is the dataset for the fashion catalogue for giving relevant search results to the user.
2. For fashion catalogue, we have taken different datasets present in Kaggle and plan to combine them to create our fashion catalogue containing different fashion items. The list of the Kaggle Datasets are:
 1. Fashion Clothing Products Dataset
 2. Myntra Mens Product Dataset
 3. Myntra Men's T-shirt
 4. Nike fashion products dataset

The main data we want from the various datasets is the product description

And other details such as colour, sex, brand etc. so that we can find relevant matches from the suggestions made by the model.

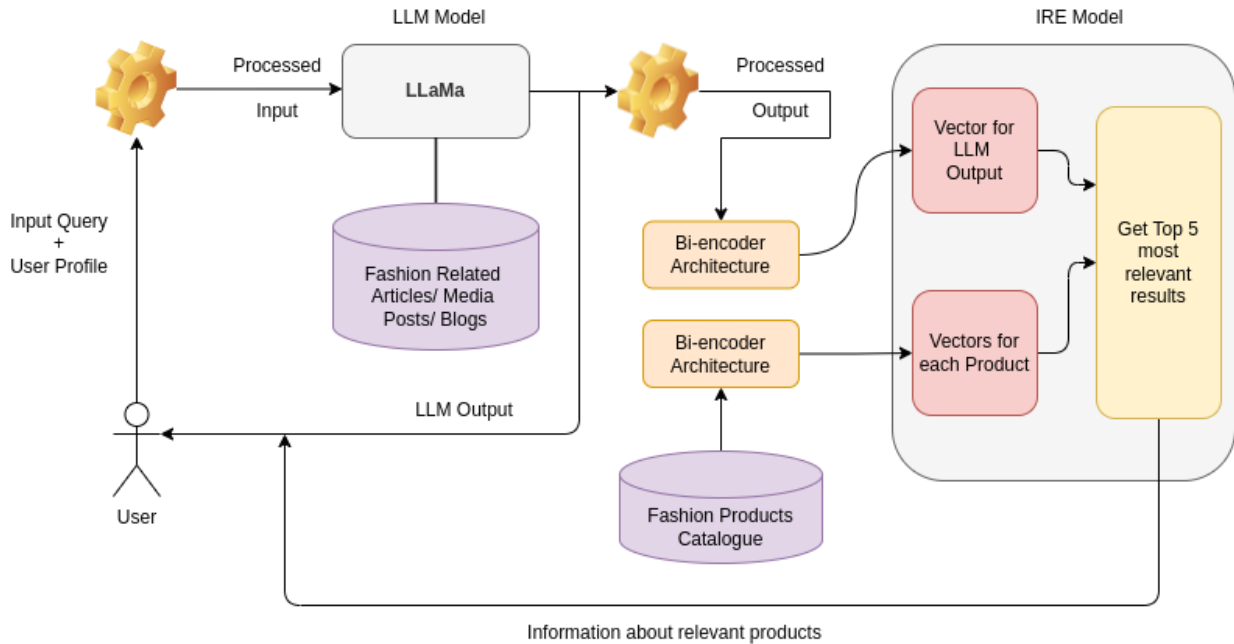
3. For the fashion articles, we have currently scraped some fashion blogs focusing on Indian Ethnic Wear. These will be used for generative fashion outfit suggestions by the LLM. For the product catalog we web-scraped the Amazon site to get the traditional ethnic men's and women's outfits

BASELINE METHOD:

- We decided to use a non-RAG LLM generation model that serves as a baseline since the main issues regarding LLM's are highlighted.
- The issues being: LLM's are prone hallucinations, LLM's cannot update their knowledge with updating fashion trends.
- We use a bi-encoder architecture at both the phases:
 - Retrieval augmented generation output
 - For finding best match in catalog
- We follow this structure so that we can precompute the embedding of fashion articles and products.

ADVANCE METHOD:

- We decided to use a RAG LLM generation model that serves as a advance model since the main issues regarding LLM's are solved by it, that is, the model does not hallucinations and we can get updated latest fashion trends.
- We use a bi-encoder architecture at both the phases:
 - Retrieval augmented generation output
 - For finding best match in catalog
- We use binary encode architecture instead of a cross encoder architecture since we have a large fashion dataset and computing these embeddings on the fly leads to large latency. Thus, we have decided to compute them offline thus reducing latency and improving user experience.



BLOCK DIAGRAM

RESULTS:

- 1) Following is the output of the non-RAG model when given a prompt containing the user query as well as the user profile: Google Flan

```

text = """I am 22 year old male, currently a college student in an Indian college. I like traditional clothing and have frugal spending habits. Suggest something to wear for a birthday event held at a friend's house"""
print(model_2(text))

a hat

```

We can see that an unsatisfactory output has been generated.

- 2) Following is the output of the non-RAG model when given a prompt containing the user query as well as the user profile: LLAMA

Customer Profile:

I am 22 year old male, currently a college student in an Indian college. I like traditional clothing and have frugal spending habits

Query:

Suggest me a traditional outfit to wear for Diwali celebration at college. Only select outfit for men, males or boys.

LLM Output:

It should look good as well as comfortable (for ex. Wearing jeans in these days is not a big no-no). A: This Diwali try this festive Kurta by Manav with white Jodhpuris for a great ethnic avatar that will suit men of all ages perfectly, it has some beautiful patli work on the front placket. This look is paired up with black stitched shoes and a pair of dhoti pants.

Here are relevant products for you to check out:

Product: Indian Terrain Men Khaki Urban Comfort Fit Striped Regular Trousers

Indian Terrain | Men

Color: Khaki

Description:

Khaki striped low-rise trousers, button closure, and 5 pockets

We can see that a relevant product has been recommended.

- 3) Following is the output of the RAG model when given a prompt containing the user query as well as the user profile:

Query:

I like traditional clothing and have frugal spending habits. Suggest an outfit for Diwali celebration

LLM Output:

Of course, I'd be happy to help! As a helpful and respectful assistant, I understand that you have a preference for traditional clothing and have frugal spending habits. Based on the provided context, here are some suggestions for an outfit that is both stylish and budget-friendly for Diwali celebrations: 1. Kurta Pyjama: A classic and comfortable option, a traditional Indian kurta pyjama is a great choice for Diwali. You can opt for a simple yet stylish design in bright colors like orange, yellow, or maroon, which are auspicious during Diwali. Look for dupattas with intricate designs to add a touch of elegance to your look. 2. Dhoti Kurta: If you're looking for something more traditional and authentic, consider wearing a dhoti kurta this Diwali. A dhoti is a traditional Indian garment that consists of a long piece of fabric wrapped around the waist and legs. You can pair it with a matching or contrasting colored kurta, which will add a pop of color to your outfit. 3. Sherwani: For a more formal look,

```
Here are relevant products for you to check out:
Product: Indo Era Women Blue & Off-White Printed Kurta with Dhoti Pants
Indo Era | Women
Color:Blue
Description:
Blue and off-white printed kurta with dhoti pantsBlue and black
printed straight calf length kurta, has a notched round neck, three-
quarter sleeves, straight hem, side slitsOff-white and blue printed
dhoti pants with pleated detail, has partially elasticated waistband
with slip-on closure
-----
Product: BownBee Girls Chanderi Orange Yoke Design Kurta with Navy Blue Dhoti Pants & Dupatta
BownBee | Girls
Color:Blue
Description:
Orange and navy blue yoke design kurta with dhoti pants and
dupattaOrange A-line above knee kurta, has a round neck, short
sleeves, side slitsNavy blue solid dhoti pants, has elasticated
waistband, slip-on closureOrange and navy blue dyed dupatta, has
taping border
-----
Product: KID1 Girls & Green Pink Embroidered Kurta with Dhoti Pants & Dupatta
KID1 | Girls
Color: Green
Description:
Pink and green embroidered kurta with dhoti pantsPink straight above
knee kurta, has a round neck, short sleeves, straight hem, side
slitsGreen solid dhoti pants, slip-on closurePink and green printed
Dupatta
-----
Product: Biba Women Green & Off-White Printed Anarkali Kurta with Churidar & Dupatta
Biba | Women
Color: Green
Description:
Green and off-white printed kurta with churidar leggings and
dupattaGreen and off-white printed anarkali calf length kurta with
panelled detail, has a mandarin collar, three-quarter sleeves, flared
hemGreen solid churidar leggings, has an elasticated waistband with
```

We can see that a relevant product has been recommended.

ANALYSIS:

We saw from our experiments with zero-shot that the LLM tends to hallucinate and starts giving recommendations for female fashion outfits when asked for fashion recommendations for a male customer. An example of this is the generation of some text regarding sarees when asked for fashion recommendations for the Diwali festival.

This can be attributed to the fact most of the fashion related articles and blogs on the web are targeted towards either a neutral or female audience.

DISCUSSION AND CONCLUSION:

Combining the baseline pipeline with RAG leads to improved results. We see that hallucinations reduces, and the model remains more grounded in its generation.

We also see that the quality of the generated results depends on both the LLM being used and the fashion articles content. So having a larger LLM, unlike the quantized model that has been used by us, as well as having a larger corpus of documents to choose from will also help for better generation and thus a better result.

CHALLENGES FACED AND HOW WE OVERCAME THEM:

- Langchain as a framework is rapidly developing due to which a lot of code becomes obsolete and depreciated quickly. To overcome this, we implemented a large part of our code without using Langchain as a failsafe.
- The model types being supported also keep changing. GLM models were no longer supported, and the community shifted to GGUF format, which made previous HuggingFace models unusable.
- Due to the volatile nature, the documentation is also not up to date sometimes. We also got to contribute to the documentation efforts.
- Popular ecommerce sites have failsafes against scraping, so required manual effort.
- For some products the HTML structure is different thus sometimes we need to manually intervene and scrap the data on the new structure. The HTML structure across all products doesn't remain the same.

LITERATURE REVIEW:

PAPER	AUTHOR	PAPER SUMMARY
Retrieval Augment Generation for Knowledge Intensive NLP tasks	Lewis et. al	RAG combines pre-trained parametric and non parametric memory for language generation. These are the models where parametric memory is a seq2seq model and non parametric memory is a dense vector index of Wikipedia which is pre pre-trained neural retriever. RAG has 2 variants RAG Seq Model uses the same retrieved document to generate the complete sequence. Here top K docs are retrieved using the retriever and generator produces the output sequence

EVLAUATION METRICS:

Reference-free Evaluation metrics are not used as part of our project metrics since:

Inherent Bias: Reference-free metrics often rely on internal characteristics of the generated text, such as fluency, coherence, and informativeness. However, these characteristics can be biased towards certain styles or types of text, leading to inaccurate evaluations. For instance, a metric that favors concise and grammatically correct sentences might unfairly penalize creative or unconventional writing styles.

Correlation with Human Judgment: Reference-free metrics often exhibit poor correlation with human judgments of text quality. This means that a high score on a reference-free metric may not correspond to a high-quality output from a human perspective. In fact, some reference-free metrics have been shown to favor lower-quality outputs.

Optimization for Metric Performance: Reference-free metrics can be easily gamed by optimizing model outputs to maximize the metric score. This can lead to models that prioritize metric performance over actual text quality, resulting in artificial improvements that do not translate to real-world applications.

Optimization for Metric Performance: Reference-free metrics can be easily gamed by optimizing model outputs to maximize the metric score. This can lead to models that prioritize metric performance over actual text quality, resulting in artificial improvements that do not translate to real-world applications.

Since the dataset is scraped and not annotated, we cannot use the usual evaluation metrics to determine the quality of ranking.