

AI-powered product search system for an e-commerce website

By Team Code Wizards

Second Phase evaluation Report

Team Leader: Yash Agrawal

Team Members: Parth Madan, Pallav Sharma, Nitin Kumar Das, Keshav Garg

Project Description: Our project aims to improve search query for e-commerce website by incorporating semantic search rather than relying on old school key word search method, which improves customers retention time which grow online revenue, increases average order value and allows personalization pipelined for specific customer's need. Even googles journey since the bought semantic database Freebase as the basis for the knowledge graph in 2010 to their introduction of MUM is quite remarkable.

Work Done till 1st phase:

1. We have created the interface for the search engine using a dummy API of Wikipedia. It extracts the data from the Wikipedia API and results are shown to the users in a very interactive manner.
2. We extracted a part of the flip kart dataset and done data preprocessing on it which includes NLP methods. These methods include following steps to perform:
 - a. Tokenization:
 - i. Split the text into sentences using sentence segmentation
 - ii. Split sentences into words using word tokenization
 - b. Text Normalization:
 - i. Convert text to lowercase
 - ii. Remove punctuation, special characters, and stop words.
 - iii. Perform stemming or lemmatization to convert words to their base form
 - c. Parts-of-speech Tagging:
 - i. Tag each word in a sentence with its grammatical role (e.g., noun, verb, adjective, etc.)
3. In the next step we embed the data using the Hugging face transformers library. We followed the following steps:
 - a. We install the library and loaded a pre-trained model.
 - b. Then we pass the tokenized text to the model to get the embeddings as the output.

- c. We now extracted the embeddings.
4. We have done the work till extracting these embeddings.

Work Done between 1st phase and 2nd phase:

1. We developed vector encodings using sentence transformers and restricted it to 384-D.
2. Our workflow would be that user enters a search query for the product he/she wants to retrieve, our aim is that if the user does not mention specific keywords in his/her query then also related results should be displayed for that we are using cosine similarity and clustering.
3. Cosine Similarity gives us a measure to calculate closeness between the query vector and the product vector when displayed in vector space of 384 dimensions. Since related vectors are close to each other so the query vector should give the relevant product result.
4. Further we are making clusters of our product vectors using k- nearest neighbors. We are trying to optimize number of clusters in our model using the algo method. We are doing this, like if there are billion records in our database k – nearest neighbors would help us to query relevant products within milliseconds.

Now we are trying to give personalized suggested product recommendations to specific user based on their search query history. We are planning to take last 10 query searches of the user.

Conclusion: By this we will develop a fully functional AI based vector search system for the e-commerce website which is both efficient and incorporate semantic search.