Design doc

Theme: Dunzo

LastModifiedOn TIMESTAMP

https://docs.google.com/document/d/1z6akRKEMuKi7Et1G8DoITJKwB1PWkrgB5Fu8cOq1Pn0/edit?ts=5d411e0a

Objective: To create a catalog of items (food, grocery, electronics and so on) and the stores which contain those items using the receipts uploaded by the riders.

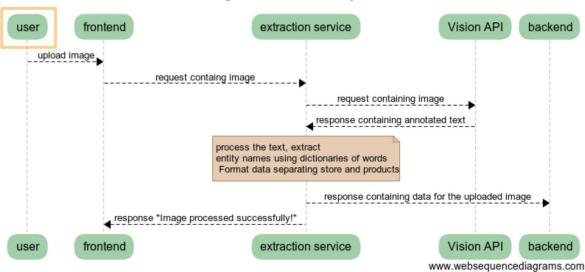
Solution: we propose a micro service architecture with 3 distinct services: text extraction service, backend service to talk to the database and a front end to take user inputs

Database schema

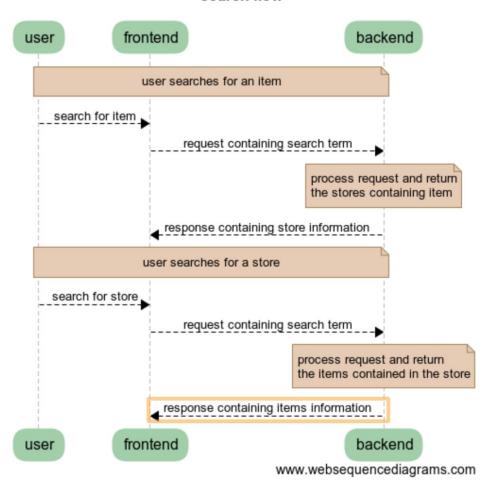
item StoreID VARCHAR(50) ltemCode VARCHAR(100) Name VARCHAR(200) Description TINYTEXT Address MEDIUMTEXT LastModifiedOn TIMESTAMP store_attributes ◇ VATID VARCHAR(45) item_type VARCHAR(100) Phone VARCHAR(45) StoreID VARCHAR(50) > Fax VARCHAR(45) AttributeKey VARCHAR(100) EMail VARCHAR(100) AttributeValue VARCHAR(100) ____ transactionlines HomePage VARCHAR(100) ○ City VARCHAR(45) EntryType VARCHAR(20) PostCode VARCHAR(10) LineNo INT(10) CountryCode VARCHAR(45) ReceiptNo VARCHAR(20) ItemCode VARCHAR(20) Description VARCHAR(200) UnitofMeasure VARCHAR(20) StoreID VARCHAR(50) StoreName VARCHAR(100) wmsentry Price DECIMAL(20,5) PEntryNo BIGINT(20) Quantity DECIMAL(20,5) P DocumentNo VARCHAR(100) DiscountAmount DECIMAL(20.5) DocumentDate DATE NetAmount DECIMAL(20.5) EntryType VARCHAR(45) VATAmount DECIMAL(20,5) ItemCode VARCHAR(100) Amount DECIMAL(20,5) Quantity DECIMAL(20,5) TotalDiscAmount DECIMAL(10,5) BaseUOM VARCHAR(20) TransDate DATE LastModifiedOn TIMESTAMP TransTime TIME imageURL MEDIUMTEXT

Flow diagram

image data extraction/entry flow



search flow



Text extraction service

This will be a simple service created using flask and we'll expose REST APIs which will consume the image and output the text in the image in the following format:

The approach to extract sentences/words from the image are as follows:

- Sharpen the image if it can happen
- Then check if the image is too blurry (80% threshold)
- If the image is too blurry, there is no point processing it so throw an error
- If the image is not blurry, get the text from the image
- Once the text is extracted, try to do an auto complete or auto correct (use google API or a dictionary of words)
- Problem: how do you determine what is the store name and what is the product name?
- Extract the text and format it accordingly