

Problem Statement : Find images for a given text ordered by relevance.

Input : Text in a text box in browser

Output : 100's of images in browser

Steps 1 : Data Collection <https://storage.googleapis.com/openimages/web/index.html>

Data set contains Image URL ,Image ID , Word , Word ID .

Create relational database in tabular format so that all (words , image_id ,image_url , word_id) can be in a single table .

Step2 : When the above table has been created then find a pre trained word vector model . Use pre-trained gensim or fastest wikipedia model . Its available free on the internet .

Step3 : Understand word similarity , read about gensim and fast text .

Step 4: When comfortable with Step 3 , then find word vectors of all the words obtained in Step1 and save all of those (word , word vector) in a dictionary /map.

Step 5 : Now take a word from the dictionary created in Step 4 and find similar 100 words .

Step 6: Sort result of Step 5 in descending order of word similarity.

Step 7 : Display images for the word in Step 6 .

Note1 : Images are not to be downloaded in local computer.

Note 2: Download csv/tsv/excel files from the source .

Note 3: For demonstration purposes a GUI is required which takes an input word and based on output renders a HTML page like an image gallery and images shown in the pages should follow word similarity obtained in step 5 .

Note 4: This project can be done in many languages but preferred language will be python as inbuilt library will help us a lot .

Challenge : Word similarity can be found very fast for 100's of word but how it can be found for millions in 20-30 seconds .

Students are advised to read the following topics :

Python

Python-flask

ZMQ

Python restful API
Python Numpy and Pandas

Data Science

Cosine Similarity

Should be able to understand cosine similarity using pen and paper .

Read basic vector algebra and linear algebra .

Interface for demonstration

Input

Enter the word for which you want to search images

Search Word:

Image Size:

Search Language:

English ☐ All Languages ☐

Submit

Output

[Search Another Word](#)

translated_word: iphone

