

1) Created dataset explanation

We have used the **figureqa-train1-v1.tar.gz(2.12 GB)** from the following link:
<https://msropendata.com/datasets/85596452-0fe3-4335-bc00-ae83ee8ffcfd>

Entry_{l} is just the entry under which rest of the attributes fall.

Image_index is the name of the image from the mentioned dataset

Graph_type is one of the five graph types used in FigureQA dataset

Passage: It is the generated passage according the image

Question: It asks about the added entity

Answer: Provides the true answer in a sentence form. It can be changed to just numeric form.

2) Contribution

I undertook the task of adding a new entity based on the images and manipulating and creating data around that added entity. I started with creating a script to extract all the names of colors that appear in FigureQA dataset (this script is not included with the submission). Then for given number of iterations, I run a loop which checks for the type of graph first. Based on the type, a predefined template generates the passage, question and from the annotations of the FigureQA dataset, Answers are calculated. All this data is added to a list which is then dumped to a json file.

3) Instructions to run code

There are no special packages required to run the code. The code was written in python 3. Simply run the file from the terminal. Only the annotations file of the original dataset is required to be in the same directory.