# SMART MECCA

Abstract

Recently, caption generation has raised a huge interest in images and videos. However, it is challenging for the models to select proper subjects in a complex background and generate desired captions in high-level vision tasks. In Artificial Intelligence (AI), the contents of an image are generated automatically which involves computer vision and NLP (Natural Language Processing).

The goal of this project is to create a smart chatbot that helps pilgrims to make their journey as easy and smooth as possible by preforming two main tasks: Answering common questions, they might come across during Hajj and Umrah. Describing images to text within the area of Holy Almasjid alharam, so people who cannot see will be able to know what inside image. In addition, we improved a program that convert caption text into voice.

The biggest challenge is most definitely being able to create a description that must capture not only the objects contained in an image, but also express how these objects relate to each other. Consider the following Image from our dataset:-
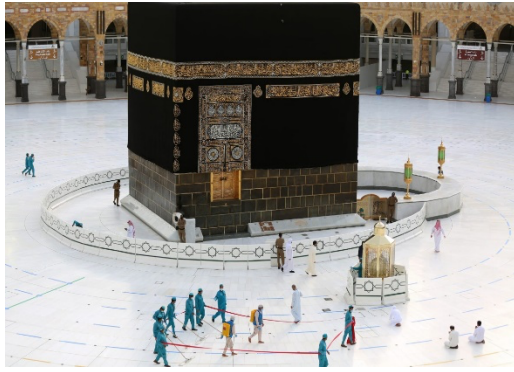
What do you see in the above image?

"The Kaaba and alsahan", "Al Kaabah","Al kaabah and Maqam Ibrahim", " Al Kaaba Alsharefh", " qeblat Al muslim".

Data Description:

In this project we created our dataset for both "chatbot" and "image caption". Merging our dataset" image caption" with "Flickr30k" datasets. Flickr30k contains 158,915 rows and 3 columns. The total number of unique images for Flickr30k are 31,783 each paired with five different captions.

Tools

- Jupyter Notebook
- Python Libraries:

  - Pandas
  - NumPy
  - Keras
  - TensorFlow
  - Sklearn
  - Matplot
  - Flask
  - Pickle
  - PIL
  - collections, random, re
  - Convolutional Neural Networks
  - seaborn
  - pydotplus
  - googletrans
  - Gtts

Algorithms

1. Exploratory Data Analysis

- Reading the data
- Cleaning the data
- Reshape the dataset and make image name as index.

2. Natural Langauge Processing

- Performed text cleaning
- Removed digits from texts.
- Made all text lowercase.
- Removed all punctuations.
- 'startseq' was added at the beginning of each caption
- 'endseq' was added at the end of each caption

3. Deep Learning

- Transfer Learning(InceptionV3):
using pre-trained InceptionV3 model, images was converted into (2048,1) vectors using the bottleneck features.
- Word Embeddings (GloVe):
using pre-trained GloVe model, common word vocabulary of captions was mapped to 200D long word vectos.
- Final Model Architecture:
Since the input consists of two parts, an **image vector** and a **captions**, we cannot use the Sequential API provided by the Keras library. For this reason, we use the **Functional API** which allows us to create Merge Models (the model visualized in appendix **A**).
The model implemented as following:
- The images and words vectors was used as an input
- Embedding and LSTM layers was applied on words vectors (RNN)
- **256D** was result from the previous layers
- Feed Forward Network ending with **2732** softmax layer to predict the next word.
4. Topic modeling and testing 'chatbot':
    - NMF& LSA
    - LDA
    - BERTopic

Get Fatwa

Question:

هل في تحية المسجد بالحرم؟

PREDICT

Answer:

نعم يوجد ركعتين تحية مسجد ، وإذا أتيت وطفت كفاك عن
تحية المسجد وإلا تصلي ركعتين أو ما تيسر ، لكن لا يلزمك
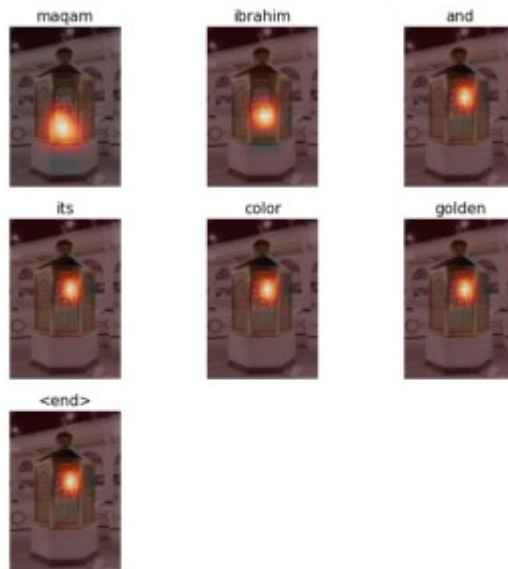في كل دخول طواف ، لو طافت الناس كلها ما خلا المطاف

Choose File   No file chosen   Submit

الوصف : مقام ابراهيم

```
BELU score: 0
Real Caption: <start>Maqam Ibrahim
Prediction Caption: maqam ibrahim and its color golden
```

Conclusion

The project successfully achieve it's goal using **Deep Learning** and **NLP**.

- Create a model that predict image description (caption).

- Improved a program that convert caption text into voice (Arabic&English)

- Arabic Fatawa Q&A.

Future Work

- Combine that two project in one Mobile app.
- Support different language.
- Callbot.

\