Abstract

In current digital world Content based Image retrieval is becoming critical problem as size of data on Internet increasing rapidly. When the image is embedded in news article it is retrieved by manipulating words annotated to that image, text placed surrounding to that image etc. Many times this annotation, caption generation is done manually. It reduces accuracy, increases time span and makes it as tough task. We proposed a new approach for generating caption for such images. Approach presented here focuses on important terms occurring in news like named entities, using term weighting find out weighted terms which helps in describing news. On other hand by image processing we find out who's in picture as it helps in making accurate caption by using face recognition and it will increase image retrieval. Some of experiments presented here shows performance of face recognition algorithms on standard datasets and also on own developed face dataset, also we train NER model on Indian names which gives better results. As it covers text and image content it helps in generating better caption and also for improving image retrieval accuracy.


- http://nlp.stanford.edu/software/CRF-ER.shtml, 15/05/2014
- https://gate.ac.uk/, 10/052014
- https://opennlp.apache.org/, 20/05/2014

**Index Terms**

Computer Science  
Image Processing

**Keywords**

Caption generation  
Name entity recognition  
Text Processing  
Face Recognition.