

Project Proposal: Open Images Dataset

Background

A semester-long project to implement big data techniques with a Google image set and prebuilt machine learning algorithm. The dataset for this project is the Google image dataset, which is consisting of ~9 million URLs to images that have been annotated with labels spanning over 6000 categories. The size of the dataset allows for training a deep neural network from a prebuild (ResNet50 model) model using the UMBC Clusters.

Scope

The end result of the project is to be able to be comfortable using big data platforms like Spark and Hadoop on a large data set. We will be using the ResNet50 pre-built model to determine if an image is a Cat or not.