CS 5542 BIG DATA ANALYTICS AND APPS
PROBLEM SET-1
16176145 CHIA-HUI AMY LIN

#1. Odd Man Out

You're given an unsorted array of integers where every integer appears exactly twice, except for one integer which appears only once. Write an algorithm (in a language of your choice) that finds the integer that appears only once.

```
# Using Binary Search function to find
# the integer that appears only once
def search(array, low, high):

        # Base cases for the search
        if low > high:
                return None

        if low == high:
                return array[low]

        # Mid number for Binary Search
        mid = low + (high - low)/2

        # Determine if Mid is even or not.
        # If it's even, search on the right side; vise versa.

        if mid%2 == 0:

                if array[mid] == array[mid+1]:
                        return search(array, mid+2, high)
                else:
                        return search(array, low, mid)

        else:
                # if mid is odd
                if array[mid] == array[mid-1]:
                        return search(array, mid+1, high)
                else:
                        return search(array, low, mid-1)



# Function call for the Odd Man Out
result = search(array, 0, len(array)-1)

if result is not None:
        print "The Odd Man in this array is %d" % result
else:
        print "Result not found in the array."
```

REFERENCE: http://code.geeksforgeeks.org/index.php

#2. System Design: Big Data Application

What kind of data do you want to extract?
Real-time images of chimpanzees on the "chimp island", activity monitor, classification ID for each chimpanzee, location.

How to collect such data?
Using surveillance camera.
By Computer Vision(group tracking, optical flow) in image processing, Collective Activity Recognition, Unified framework for multi-target tracking.

How big data analytics can uncover the unexpected in your data?
Data Mining:
Big data analytics will help people examine large amounts of data and discover the patterns – activities, behaviors…etc. This information can be used for further analysis to help find potential problems on a chimpanzee (whether certain symptoms indicate it's sick), monitor activities of the chimpanzees, answer complex questions either from the crowd or from the zoo keepers. The big data analytics will also skip through the chaotic and noise in the extracting data, bullet point what's more relevant to the system or any question being asked, pick out the information to assess the most likely outcome and accelerate the speed of making decisions.

Predictive Analytics/Modeling:
The historical data can be run through statistical algorithms and machine-learning techniques to identify the likelihood of future outcomes. Furthermore, the data will be updated through real-time application to provide a best assessment in the future or even make real-time decisions.

How does the learning improve your system?
Training algorithm will adjust the weights reducing the error between the known output values and the actual values (Artificial Neural Network). -- > Machine Learning
Capable of handling multiple activities in the time frame or even real-time efficiently
Enable segmentation of individuals into different collective activities (we are able to monitor multiple chimpanzees and collect their data all at once)

REFERENCE:
http://cvgl.stanford.edu/posters/choi_cvpr11_poster.pdf
http://vhosts.eecs.umich.edu/vision//activity-dataset.html

Draw the workflow of the proposed system and explain the process of the system