

# CIS4930 Group Project

*Dax Gerts, Christine Moore, Carl Amko, Denzel Mathew, Aaron Silcott*

*December 16, 2015*

## Introduction

The following report details the

## Literature Review

W

## Overview of Data

Our working dataset was created using a python script making calls to the OMDb API.

```
## Title: IMDB sample database creator
## Author: Dax Gerts
## Date: 14 December 2016
## Description: used the OMDb API to rapidly build a clean JSON data sample of the IMDB database

import requests
import numpy as np
import json
import urllib2
import csv
import codecs

sample_size = 30000

random_ids = np.random.choice(5262000,size=sample_size,replace=False)

data = [None] * sample_size

for i in range(sample_size):
    temp_id = "%07d" % random_ids[i]
    #text = requests.get('http://www.omdbapi.com/?i=tt' + temp_id + '&plot=short&r=json').text
    data[i] = json.load(urllib2.urlopen('http://www.omdbapi.com/?i=tt' + temp_id + '&plot=short&r=json'))
    print("QUERY#" + str(i))

with codecs.open('imdb_re_30k_sample.txt','w','utf8') as f:
    json.dump(data, f)
```

## Method and Materials

## Results

## Discussion

## Conclusion

## References & Resources

The Internet Movie Database <http://www.imdb.com>

IMDbPY - Python API and Materials for IMDB searches <http://www.imdbpy.sourceforge.net>

The Open Movie Database <http://www.omdbapi.com/>