

# CS595 Assignment 3

Jon Robison

September 30, 2013

Q1. Download the 1000 URIs from assignment 2...remove formatting  
Large collection of files included in modified/ directory

Q2. Choose a query term...to ten documents...compute TFIDF values for  
the term in each of the 10 documents.

TFIDF TF IDF URI

```
.0096 .0009 10.7398 http://themoneyteam.com/
.0279 .0026 10.7398 http://mashable.com/2013/09/23/floyd-mayweather-manziel/
.0644 .0060 10.7398 http://instagram.com/p/eqTrAfFCz0/
.0032 .0003 10.7398 http://www.youtube.com/watch?v=M4EC7l_Dt30
.0461 .0043 10.7398 http://oohwonder.tumblr.com/post/61861325994/floyd-mayweather
.0397 .0037 10.7398 http://naija2day.com/2013/09/mayweather-wants-miley-cyrus-as-escort/?utm=
.1116 .0104 10.7398 http://floydmayweather.com/
.0644 .0060 10.7398 http://soundcloud.com/ronjii/floyd-mayweather-by-qilla-fang
.0193 .0018 10.7398 http://paquesuene.net/v1/raymond-y-miguel-se-cura-con-mayweather-vs-canelo-pelea-completa
.0719 .0067 10.7398 http://www.vibevixen.com/2013/09/mayweather-wants-miley-cyrus-as-next-escort/
```

Q3. Now rank the same 10 URIs from question 2, but this time by their  
PageRank.

PageRank URI

```
.4 http://themoneyteam.com
.4 http://floydmayweather.com
0 http://mashable.com/2013/09/23/floyd-mayweather-manziel
0 http://instagram.com/p/eqTrAfFCz0
0 http://www.youtube.com/watch?v=M4EC7l_Dt30
0 http://oohwonder.tumblr.com/post/61861325994/floyd-mayweather
0 http://naija2day.com/2013/09/mayweather-wants-miley-cyrus-as-escort/
0 http://soundcloud.com/ronjii/floyd-mayweather-by-qilla-fang
0 http://paquesuene.net/v1/raymond-y-miguel-se-cura-con-mayweather-vs-canelo-pelea-completa
0 http://www.vibevixen.com/2013/09/mayweather-wants-miley-cyrus-as-next-escort/
```

Briefly compare and contrast the rankings produced in questions 2 and 3.

A3: Top TFIDF was floydmayweather.com, second was themoneyteam.com. These two were also the only two non-zero scoring options, and I infer TFIDF was heavily weighted in the pagerank calculation. It is unfortunate the other pages weren't page ranked, but I predict they would have linearly small results.