

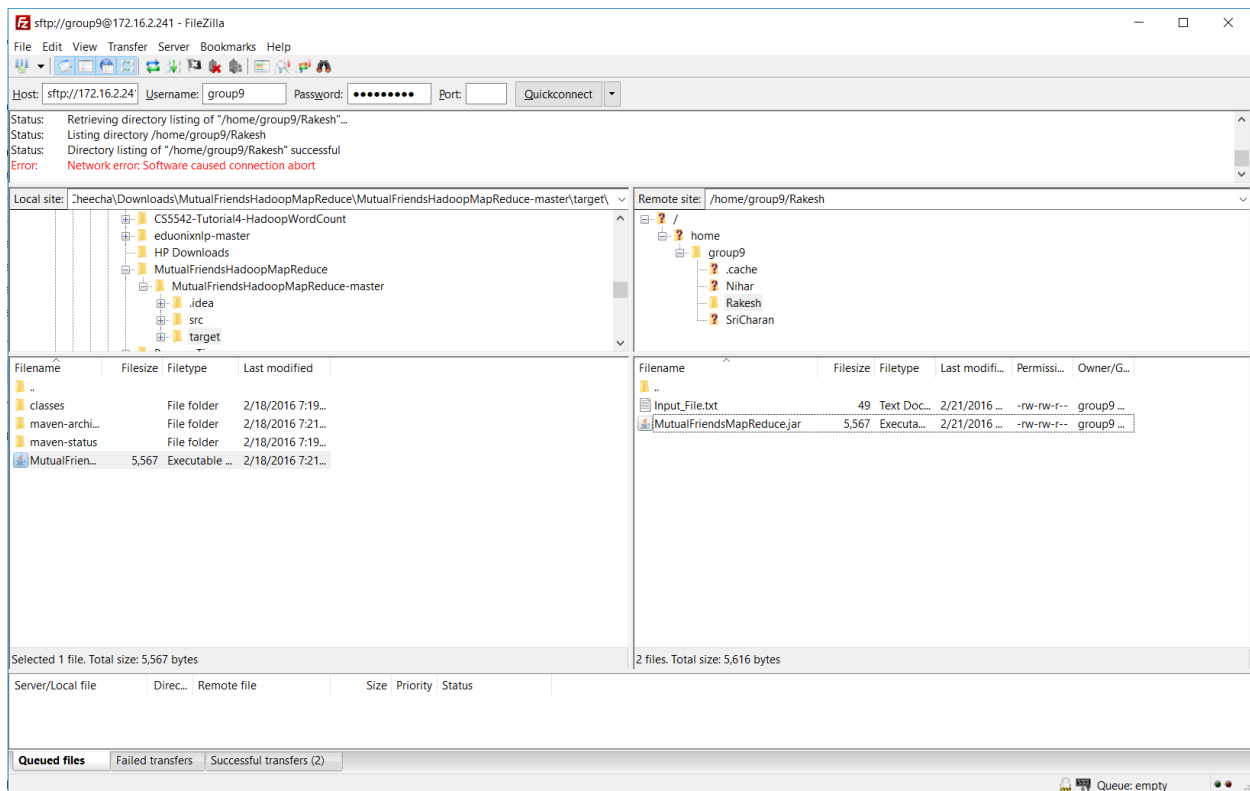
# CS5542 Big Data Apps and Analytics

## LAB ASSIGNMENT #4

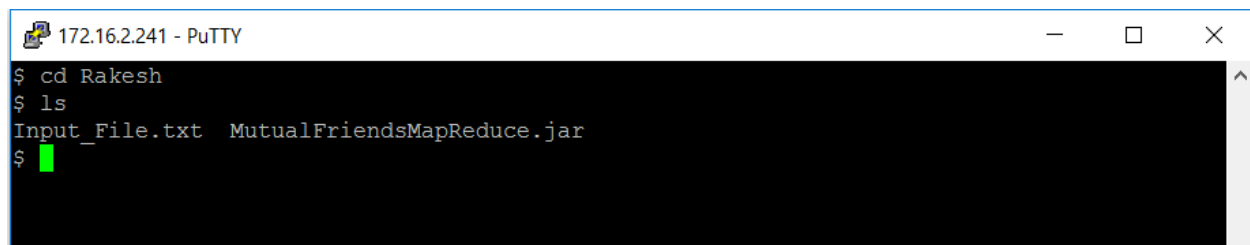
### REPORT and SCREEN SHOTS

**1.HadoopMapReduce AlgorithmImplement MapReduce algorithmforfindingFacebook common friendsproblem and run theMapReduce job on Apache Hadoop.Write a report including your algorithm and result screenshots.**

1. Push the Jar file and Input.txt file into the cloud using the FileZilla client.



2. open the putty and run the command ls to view the files in cloud



3. Create a directory in HDFS using the following command

```
172.16.2.241 - PuTTY
$ hadoop fs -mkdir Rakesh_hdfs
$
```

4. Create an input directory to move the input files into the folder

```
172.16.2.241 - PuTTY
$ ls
Input_File.txt  MutualFriendsMapReduce.jar
$ hadoop fs -mkdir Rakesh_hdfs/input
$
```

5. Move the input files such as Input\_file.txt and jar files in to the Hadoop HDFS using command `hadoop fs -put input_directory target_directory`

```
172.16.2.241 - PuTTY
$ ls
Input_File.txt  MutualFriendsMapReduce.jar
$ hadoop fs -mkdir Rakesh_hdfs/input
$ hadoop fs -put MutualFriendsMapReduce.jar Rakesh_hdfs/
$ hadoop fs -put Input_File.txt Rakesh_hdfs/input/
$ hadoop fs -ls Rakesh_hdfs/
Found 2 items
-rw-r--r--  3 group9 supergroup      5567 2016-02-22 03:02 Rakesh_hdfs/MutualFriendsMap
Reduce.jar
drwxr-xr-x  - group9 supergroup        0 2016-02-22 03:03 Rakesh_hdfs/input
$
```

6. View the files in HDFS filesystem directory

```
172.16.2.241 - PuTTY
$ ls
Input_File.txt  MutualFriendsMapReduce.jar
$ hadoop fs -mkdir Rakesh_hdfs/input
$ hadoop fs -put MutualFriendsMapReduce.jar Rakesh_hdfs/
$ hadoop fs -put Input_File.txt Rakesh_hdfs/input/
$ hadoop fs -ls Rakesh_hdfs/
Found 2 items
-rw-r--r--  3 group9 supergroup      5567 2016-02-22 03:02 Rakesh_hdfs/MutualFriendsMap
Reduce.jar
drwxr-xr-x  - group9 supergroup        0 2016-02-22 03:03 Rakesh_hdfs/input
$ hadoop fs -ls Rakesh_hdfs/input
Found 1 items
-rw-r--r--  3 group9 supergroup        49 2016-02-22 03:03 Rakesh_hdfs/input/Input_File
.txt
$
```

7. Run the Jar file using `hadoop jar /path to Jar classname input_path output_path`

```
172.16.2.241 - PuTTY
$ hadoop jar MutualFriendsMapReduce.jar MutualFriends Rakesh_hdfs/input/Input_File.txt Ra
kesh_hdfs/output
16/02/22 03:13:44 INFO client.RMProxy: Connecting to ResourceManager at KC-SCE-CS5542-1/1
72.16.2.241:8032
16/02/22 03:13:45 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing
not performed. Implement the Tool interface and execute your application with ToolRunner
to remedy this.
16/02/22 03:13:45 INFO input.FileInputFormat: Total input paths to process : 1
16/02/22 03:13:45 INFO mapreduce.JobSubmitter: number of splits:1
16/02/22 03:13:45 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_14556909157
80_0037
16/02/22 03:13:45 INFO impl.YarnClientImpl: Submitted application application_14556909157
80_0037
16/02/22 03:13:45 INFO mapreduce.Job: The url to track the job: http://KC-SCE-CS5542-1:80
88/proxy/application_1455690915780_0037/
16/02/22 03:13:45 INFO mapreduce.Job: Running job: job_1455690915780_0037
16/02/22 03:13:56 INFO mapreduce.Job: Job job_1455690915780_0037 running in uber mode : f
alse
16/02/22 03:13:56 INFO mapreduce.Job:  map 0% reduce 0%
16/02/22 03:14:05 INFO mapreduce.Job:  map 100% reduce 0%
16/02/22 03:14:15 INFO mapreduce.Job:  map 100% reduce 50%
16/02/22 03:14:22 INFO mapreduce.Job:  map 100% reduce 100%
16/02/22 03:14:22 INFO mapreduce.Job: Job job_1455690915780_0037 completed successfully
16/02/22 03:14:22 INFO mapreduce.Job: Counters: 49
    File System Counters
      FILE: Number of bytes read=172
      FILE: Number of bytes written=345085
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=186
      HDFS: Number of bytes written=57
      HDFS: Number of read operations=9
      HDFS: Number of large read operations=0
      HDFS: Number of write operations=4
    Job Counters
      Launched map tasks=1
      Launched reduce tasks=2
      Data-local map tasks=1
      Total time spent by all maps in occupied slots (ms)=6989
      Total time spent by all reduces in occupied slots (ms)=12759
      Total time spent by all map tasks (ms)=6989
      Total time spent by all reduce tasks (ms)=12759
      Total vcore-seconds taken by all map tasks=6989
      Total vcore-seconds taken by all reduce tasks=12759
      Total megabyte-seconds taken by all map tasks=7156736
      Total megabyte-seconds taken by all reduce tasks=13065216
```

```
HDFS: Number of large read operations=0
HDFS: Number of write operations=4
Job Counters
  Launched map tasks=1
  Launched reduce tasks=2
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=6989
  Total time spent by all reduces in occupied slots (ms)=12759
  Total time spent by all map tasks (ms)=6989
  Total time spent by all reduce tasks (ms)=12759
  Total vcore-seconds taken by all map tasks=6989
  Total vcore-seconds taken by all reduce tasks=12759
  Total megabyte-seconds taken by all map tasks=7156736
  Total megabyte-seconds taken by all reduce tasks=13065216
Map-Reduce Framework
  Map input records=5
  Map output records=18
  Map output bytes=138
  Map output materialized bytes=164
  Input split bytes=137
  Combine input records=0
  Combine output records=0
  Reduce input groups=9
  Reduce shuffle bytes=164
  Reduce input records=18
  Reduce output records=9
  Spilled Records=36
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=94
  CPU time spent (ms)=2900
  Physical memory (bytes) snapshot=849539072
  Virtual memory (bytes) snapshot=4145049600
  Total committed heap usage (bytes)=1052246016
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=49
File Output Format Counters
  Bytes Written=57
```

#### 8. View the files in output directory

We can see the `_SUCCESS` for successive run of job and output reduced key value pairs in `part-r-*` files

```
172.16.2.241 - PuTTY
$ hadoop fs -ls Rakesh_hdfs/output
Found 3 items
-rw-r--r--  3 group9 supergroup          0 2016-02-22 03:14 Rakesh_hdfs/output/_SUCCESS
-rw-r--r--  3 group9 supergroup      38 2016-02-22 03:14 Rakesh_hdfs/output/part-r-00000
-rw-r--r--  3 group9 supergroup      19 2016-02-22 03:14 Rakesh_hdfs/output/part-r-00001
$
```

#### 9. View the output files using cat command and we can see the Reduced key value pairs.

The common friends for the each pair can see the value of those index.

```
172.16.2.241 - PuTTY
$ hadoop fs -ls Rakesh_hdfs/output
Found 3 items
-rw-r--r--  3 group9 supergroup          0 2016-02-22 03:14 Rakesh_hdfs/output/_SUCCESS
-rw-r--r--  3 group9 supergroup      38 2016-02-22 03:14 Rakesh_hdfs/output/part-r-00000
-rw-r--r--  3 group9 supergroup      19 2016-02-22 03:14 Rakesh_hdfs/output/part-r-00001
$ hadoop fs -cat Rakesh_hdfs/output/part-r-00000
AB      CD
AD      BC
BC      ADE
BE      CD
CD      ABE
DE      BC
$ hadoop fs -cat Rakesh_hdfs/output/part-r-00001
AC      BD
BD      ACE
CE      BD
$
```

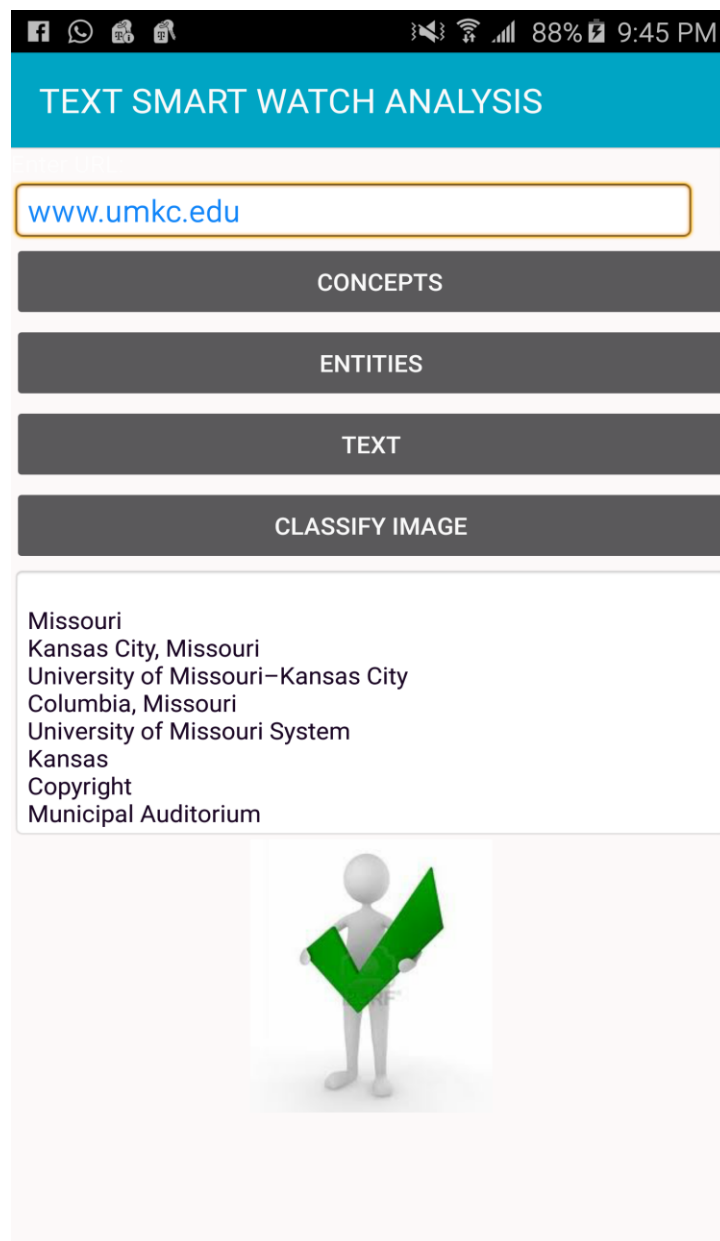
## 2. Smartphone/WatchApplicationImplement a smartwatch/smartphone application using existing speech services/image services (e.g., IBM Alchemyapi, Face++) related to your project.

I have done the SmartPhone application to perform the Text analysis and Image analysis using the Alchemy API.

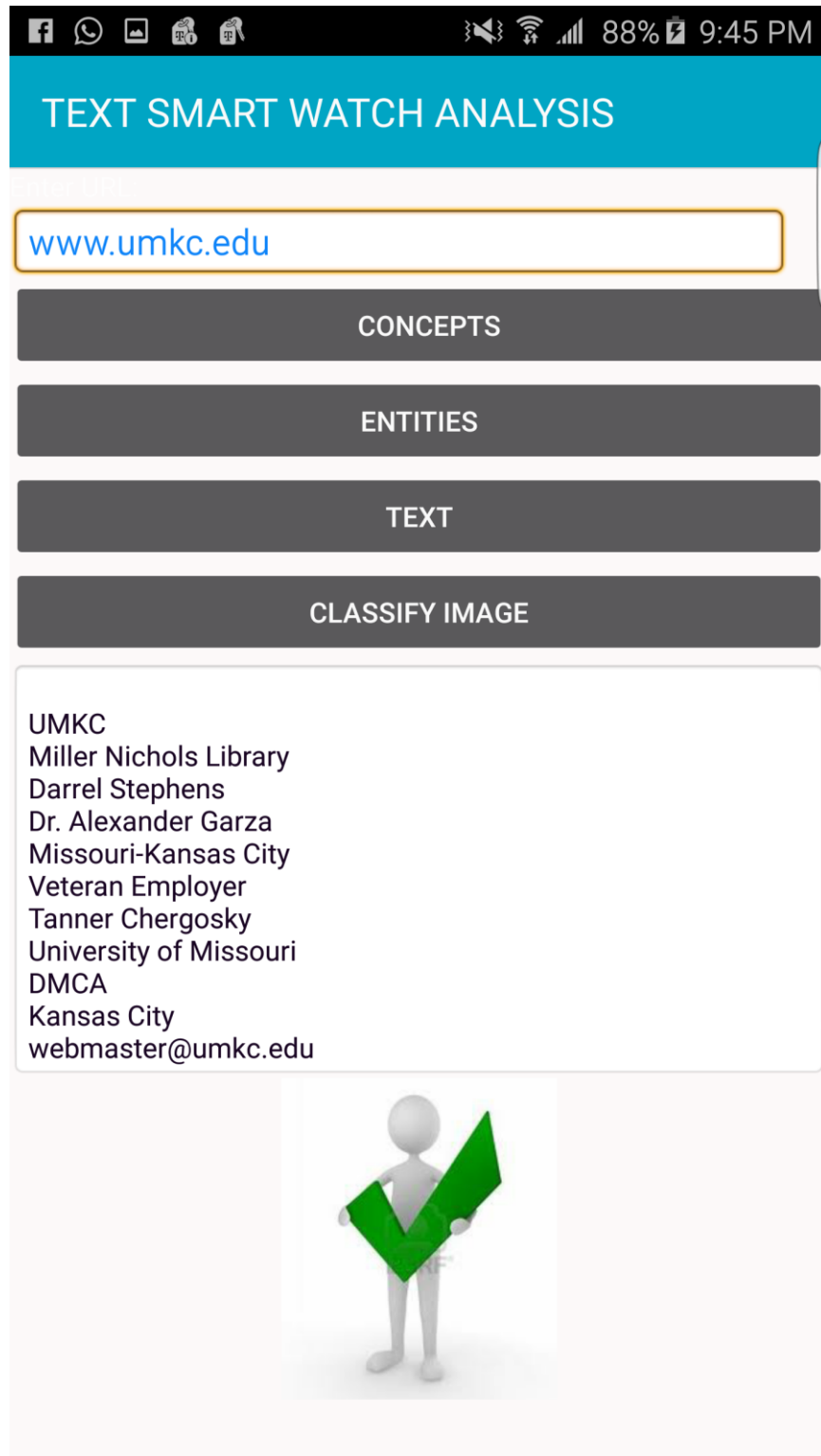
I have used the Alchemy API to perform the ENTITY, CONCEPT and TEXT analysis over the TEXT or URL provided as input.

And also used the IMAGE CLASSIFY from the Alchemy API to classify the image given as Input.

### 1. Concepts analysis of url “[www.umkc.edu](http://www.umkc.edu)” using Alchemy API



2. ENTITY analysis on the URL [www.umkc.edu](http://www.umkc.edu) provided as input



3. TEXT analysis of URL [www.umkc.edu](http://www.umkc.edu)

88% 9:45 PM

TEXT SMART WATCH ANALYSIS

[www.umkc.edu](http://www.umkc.edu)

CONCEPTS

ENTITIES

TEXT

CLASSIFY IMAGE

New mural at Miller Nichols Library pays tribute to richer learning... more >>

Darrel Stephens to receive Alumni Spotlight Award... more >>


UMKC's weekly arts, sports and culture roundup... more >>

Dr. Alexander Garza named Alumnus of the Year ... more >>

Tanner Chergosky Lives Heavy, Travels Light ... more >>

© Curators of the University of Missouri. DMCA and other copyright information. University of Missouri-Kansas City | Kansas City, MO 64110 | (816) 235-1000

UMKC is an An Equal Opportunity/Access/Affirmative Action/Pro Disabled and Veteran Employer. Email questions or comments about this website to [webmaster@umkc.edu](mailto:webmaster@umkc.edu).





#### 4. IMAGE CLASSIFICATION on the image provided

