

Lab Assignment 4

1. Hadoop MapReduce Algorithm

Implement MapReduce algorithm for finding Facebook common friends problem and run the MapReduce job on Apache Hadoop. Write a report including your algorithm and result screenshots.

Description: The Hadoop map reduce code to determine the mutual friends has been compiled and run on IntelliJ. It creates a jar file. This file is imported to cloudera through Filezilla. So in cloudera terminal Hadoop commands were run to get the desired output, which was shown below.

MutualFriendsMapReduce - [F:\BAA\MutualFriendsHadoopMapReduce] - MutualFriends.java - IntelliJ IDEA 15.0.1

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

MutualFriendsHadoopMapReduce target MutualFriendsMapReduce-1.0-SNAPSHOT.jar

Project Structure: MutualFriendsHadoopMapReduce [MutualFriendsMapReduce]

- idea
- src
 - main
 - java
 - MutualFriends
 - target
 - classes
 - maven-archiver
 - maven-status
 - MutualFriendsMapReduce-1.0-SNAPSHOT.jar
 - MutualFriendsMapReduce.iml
 - pom.xml
 - README.md

```
import ...

public class MutualFriends {

    public static class FriendsMapper
        extends Mapper<Object, Text, Text, Text> {
        private Text m_id = new Text();
        private Text m_others = new Text();

        public void map(Object key, Text value, Context context)
            throws IOException, InterruptedException {
            // In our case, the key is null and the value is one line of our input file.
            // Split by space to separate the user and its friends list.
            String line = value.toString();
            String[] split = line.split(" ");
            String subject = split[0];
            String[] friends = Arrays.copyOfRange(split, 1, split.length);
```

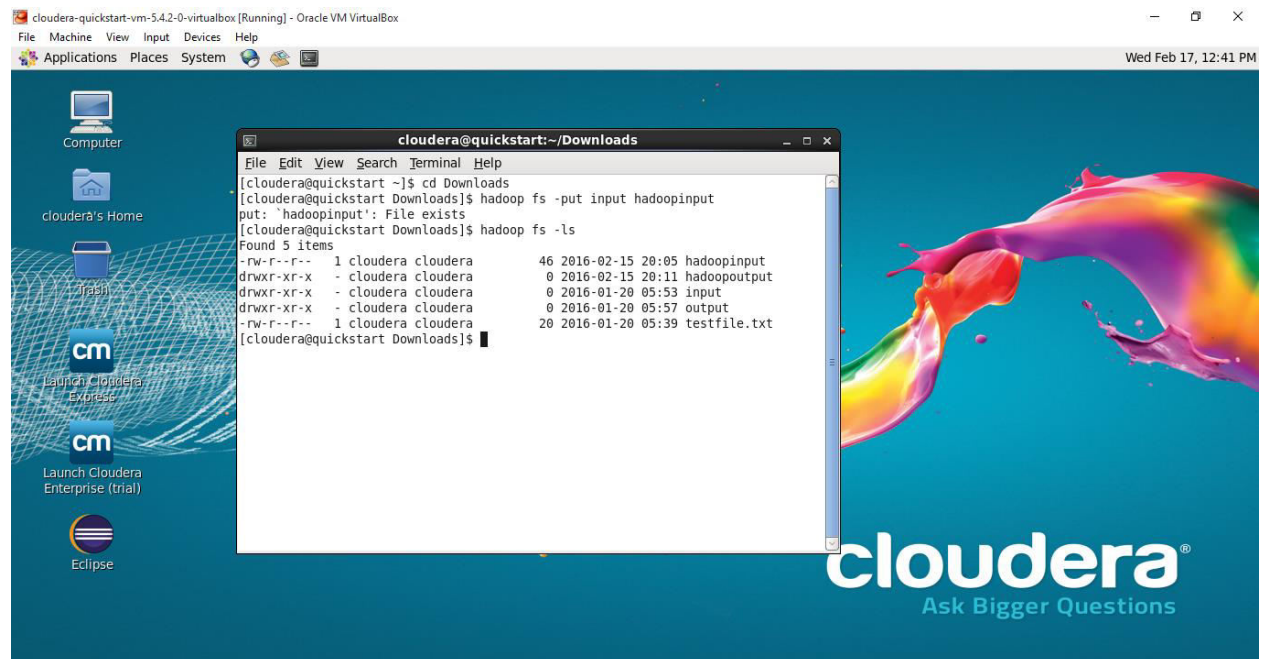
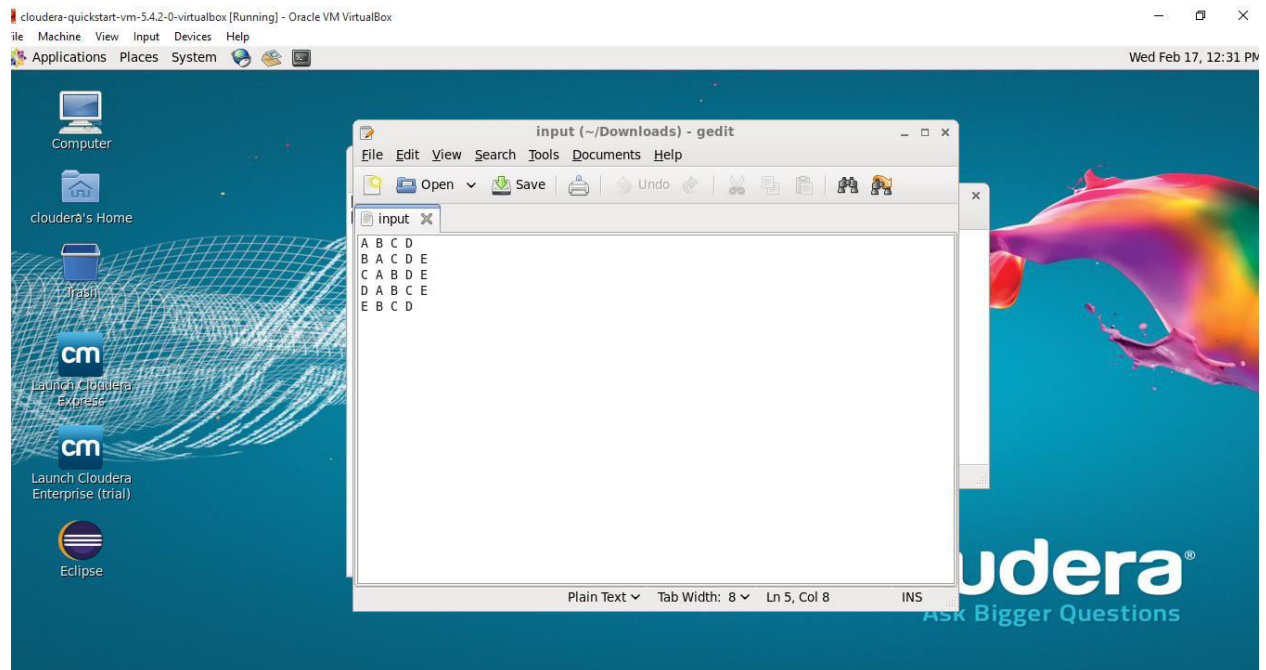
Terminal

```
[INFO] skip non existing resourceDirectory F:\BAA\MutualFriendsHadoopMapReduce\src\test\resources
[INFO] --- maven-compiler-plugin:3.1:testCompile (default-testCompile) @ MutualFriendsMapReduce ---
[INFO] No sources to compile
[INFO] --- maven-surefire-plugin:2.12.4:test (default-test) @ MutualFriendsMapReduce ---
[INFO] No tests to run.
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ MutualFriendsMapReduce ---
[INFO] Building jar: F:\BAA\MutualFriendsHadoopMapReduce\target\MutualFriendsMapReduce-1.0-SNAPSHOT.jar
[INFO] BUILD SUCCESS
[INFO] Total time: 33.003 s
[INFO] Finished at: 2016-02-17T15:43:09-06:00
```

Event Log

Error running mutualfrids: No junit.jar in Module 'MutualFriendsMapReduce' runtime scope (4 minutes ago)

Running Facebook Common Friends: Facebook has a list of friends (Note that friends are a bidirectional



```

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
[cloudera@quickstart Downloads]$ hadoop jar MutualFriendsMapReduce-1.0-SNAPSHOT.
jar MutualFriends hadoopinput hadoopoutput
16/02/17 12:56:59 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/02/17 12:57:35 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
16/02/17 12:57:52 INFO input.FileInputFormat: Total input paths to process : 1
16/02/17 12:57:53 INFO mapreduce.JobSubmitter: number of splits:1
16/02/17 12:58:02 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1455739104089_0001
16/02/17 13:00:07 INFO impl.YarnClientImpl: Submitted application application_1455739104089_0001
16/02/17 13:00:18 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1455739104089_0001/
16/02/17 13:00:18 INFO mapreduce.Job: Running job: job_1455739104089_0001
16/02/17 13:01:47 INFO mapreduce.Job: Job job_1455739104089_0001 running in uber mode : false
16/02/17 13:01:47 INFO mapreduce.Job: map 0% reduce 0%
16/02/17 13:02:55 INFO mapreduce.Job: map 100% reduce 0%
16/02/17 13:03:26 INFO mapreduce.Job: map 100% reduce 100%
16/02/17 13:03:27 INFO mapreduce.Job: Job job_1455739104089_0001 completed successfully
16/02/17 13:03:28 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=180
    FILE: Number of bytes written=220703
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=168
    HDFS: Number of bytes written=57
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=63103
    Total time spent by all reduces in occupied slots (ms)=26939
    Total time spent by all map tasks (ms)=63103

```

```

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
  Total megabyte-seconds taken by all map tasks=64617472
  Total megabyte-seconds taken by all reduce tasks=27585536
  Map-Reduce Framework
    Map input records=5
    Map output records=18
    Map output bytes=138
    Map output materialized bytes=180
    Input split bytes=122
    Combine input records=0
    Combine output records=0
    Reduce input groups=9
    Reduce shuffle bytes=180
    Reduce input records=18
    Reduce output records=9
    Spilled Records=36
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=387
    CPU time spent (ms)=3370
    Physical memory (bytes) snapshot=305479680
    Virtual memory (bytes) snapshot=3007234048
    Total committed heap usage (bytes)=164630528
  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=46
  File Output Format Counters
    Bytes Written=57
[cloudera@quickstart Downloads]$

```

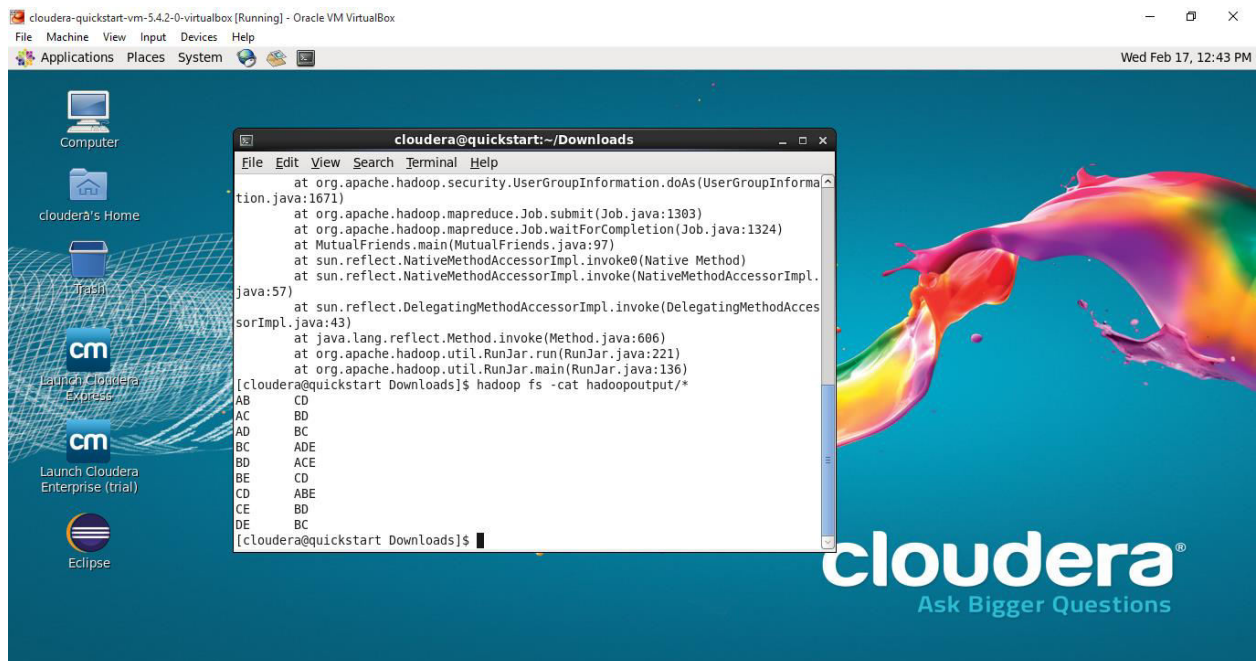
```

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
Reduce input groups=9
Reduce shuffle bytes=180
Reduce input records=18
Reduce output records=9
Spilled Records=36
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=387
CPU time spent (ms)=3370
Physical memory (bytes) snapshot=305479680
Virtual memory (bytes) snapshot=3007234048
Total committed heap usage (bytes)=164630528

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=46
File Output Format Counters
Bytes Written=57
[cloudera@quickstart Downloads]$ hadoop fs -cat hadoopoutput/*
AB      CD
AC      BD
AD      BC
BC      ADE
BD      ACE
BE      CD
CD      ABE
CE      BD
DE      BC
[cloudera@quickstart Downloads]$

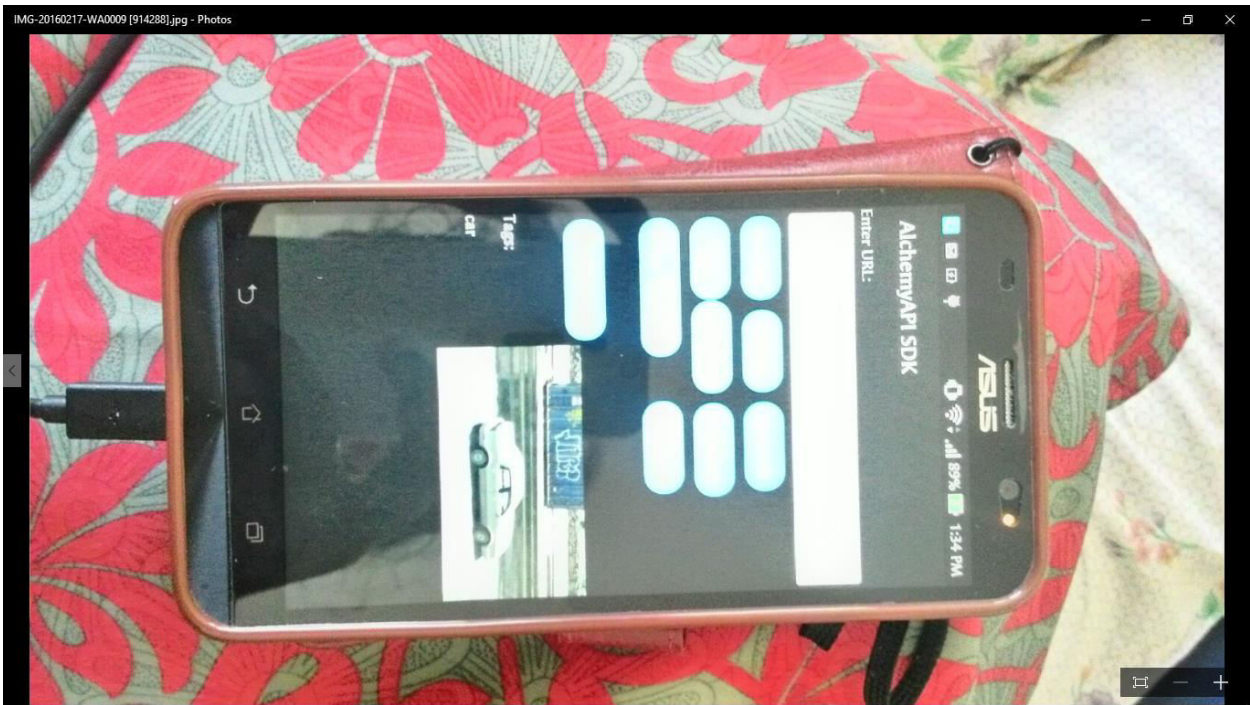
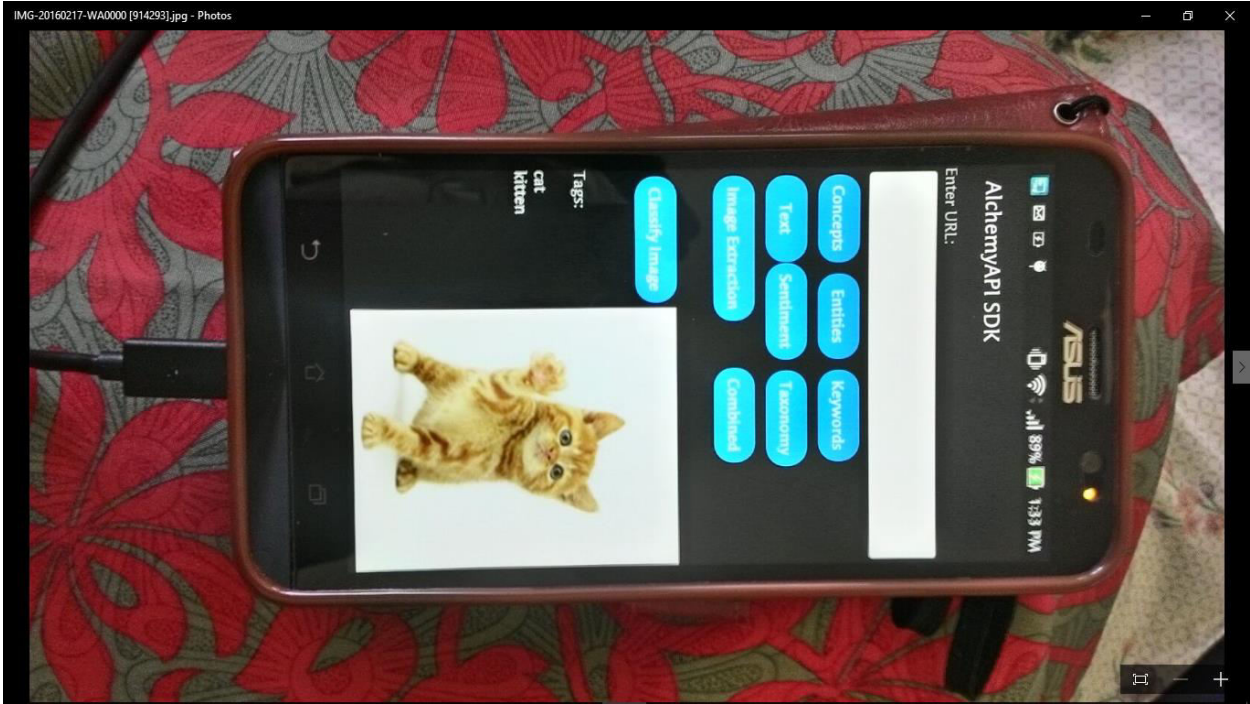
```

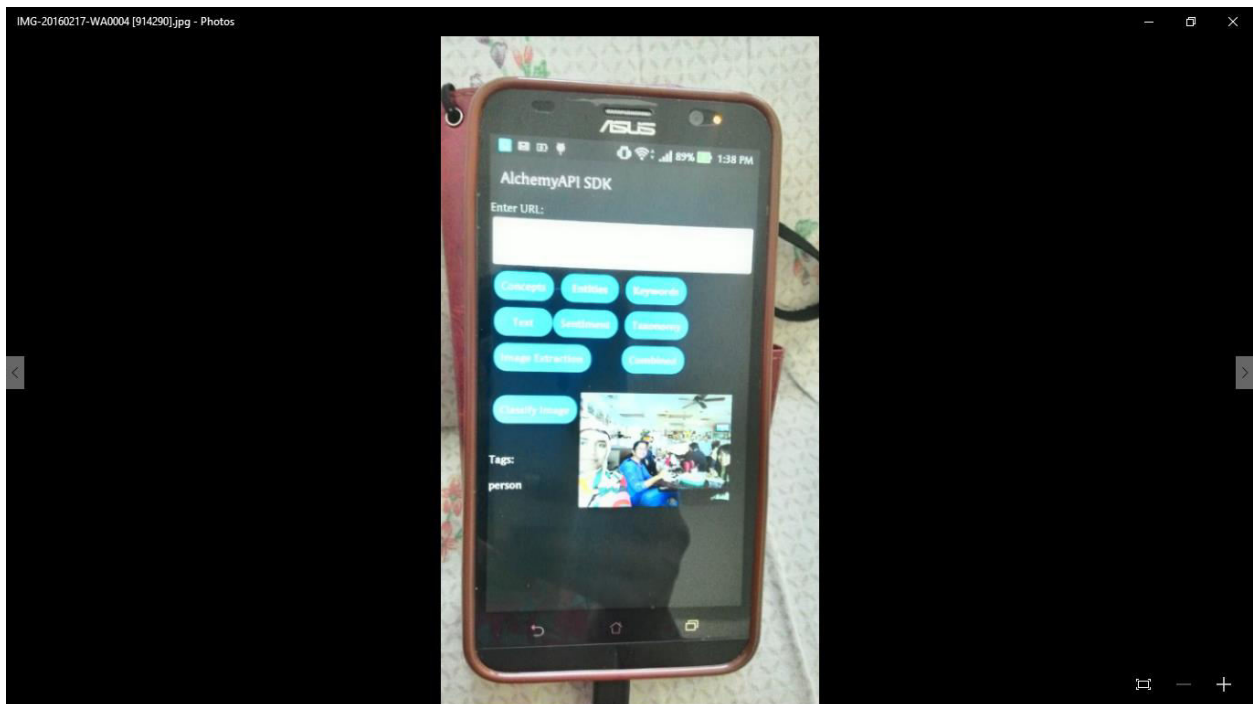
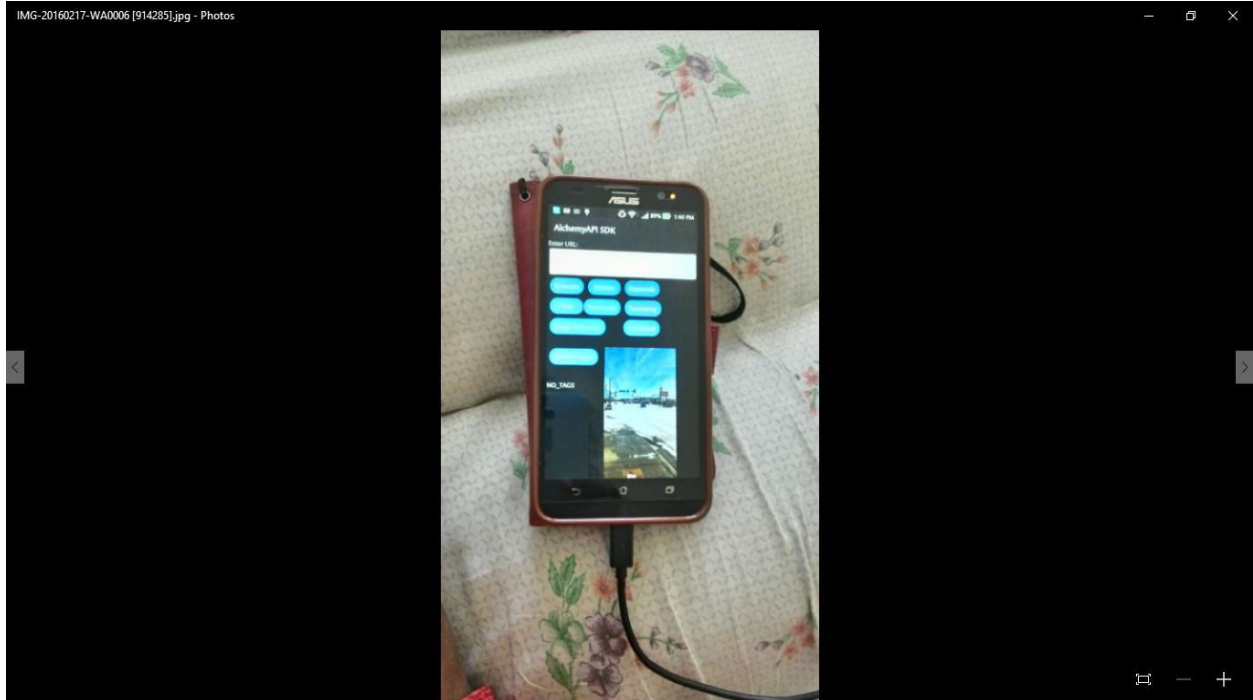


2. Smartphone/Watch Application

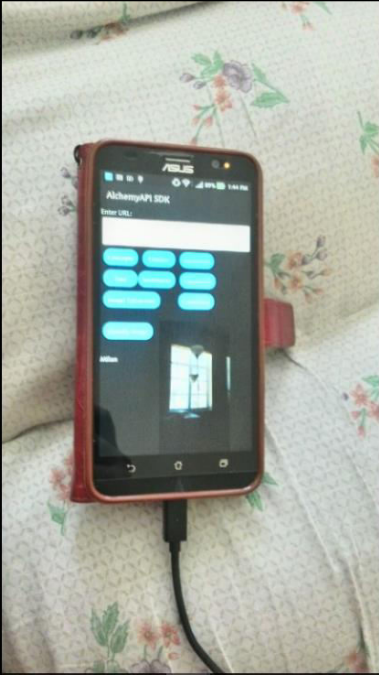
Implement a smartwatch/smartphone application using existing speech services/image services (e.g., IBM Alchemyapi, Face++) related to your project.

Description: This application uses IBM alchemy API. What it does is, it takes the image and tell us what it is. The results are shown below.





IMG-20160217-WA0007 [914289].jpg - Photos



IMG-20160217-WA0008 [914286].jpg - Photos

