|  |
| --- |
| Cairo University Multimedia  Faculty of Engineering Spring 2017  Computer Department |

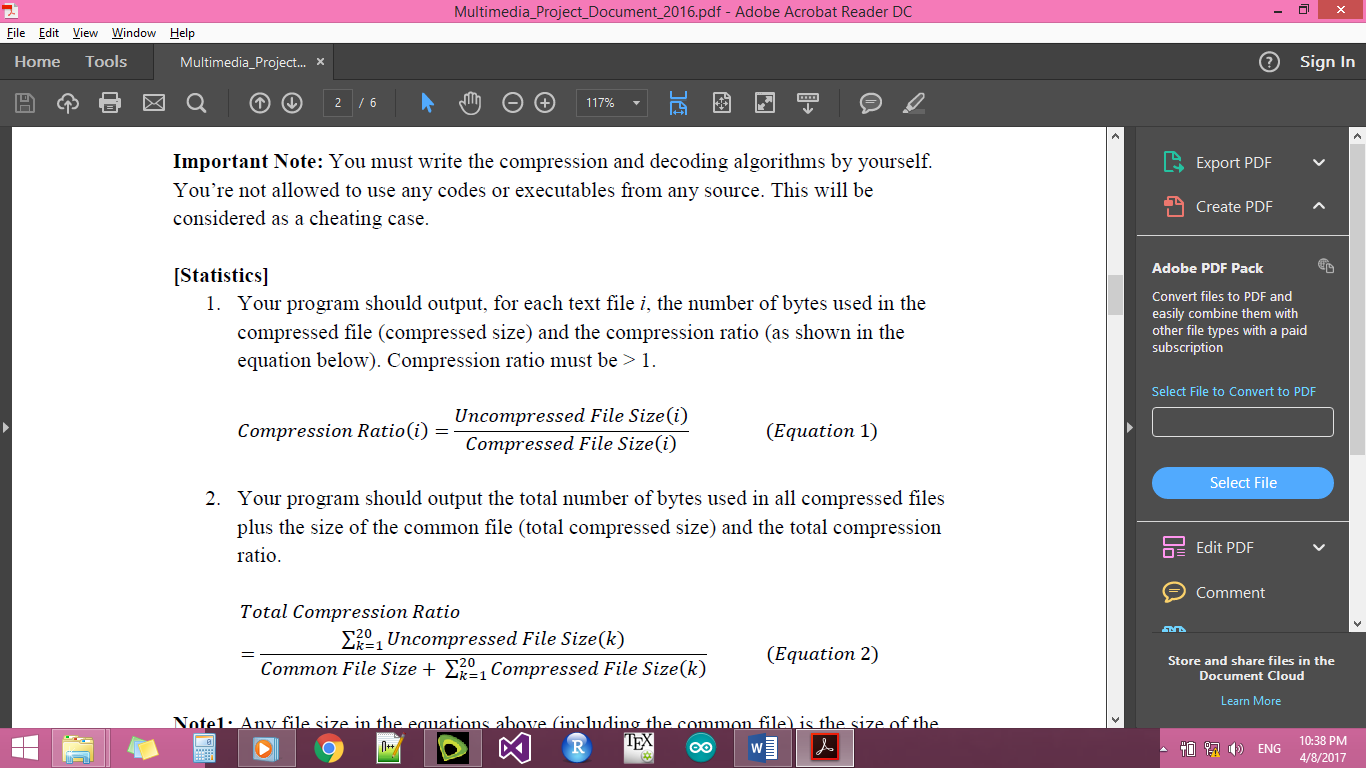
Multimedia Project Requirement Document

# Description

* The project is to find the best lossless compression technique that can be used to compress images of text documents.
* The test data (found in the attached data set folder) that we use in the project is 50 “jpg” images contains text. These images will be converted to binary level images (i.e. it will be represented in terms of zeros and ones only).
* You’re required to make any needed data analysis (e.g. calculate probabilities …etc.), compress the binary images and decode them correctly to generate the original files.
* The project is a competition between teams. The winner is the one that gets the smallest number of bytes used in the encoded message (highest compression ratio).

# Requirements

* Compression algorithm :Find the best possible lossless compression method, by best we mean less number of generated bytes
* Encoder: Takes any input image and constructs a lossless binary compressed file. Your program must write one binary compressed file for each input image ,This file must contain any extra information if needed in the decoding process of this image (e.g. number of characters in this file …etc.)
* Decoder: Takes a compressed file and decodes it to generate the original file.
* Your program should output, for each text file *i*, the number of bytes used in the compressed file (compressed size) and the compression ratio (as shown in the equation below). Compression ratio must be > 1



* Your program should output the total number of bytes used in all compressed files plus the size of the common file (total compressed size) and the total compression ratio.

# Conditions any violation to these conditions will cause loss of marks

* You are not allowed to save any information specific for one of the test files directly in the decoder, instead save them in the compressed file, then read the compressed file in the decoder that is because we use this compressed file for evaluation.
* Compress each image to a separated file.

# Evaluation criteria

* Correct lossless compression (20%).
* Correct lossless decompression (20%).
* Project document and individual rating (10%).
* Competition and innovation ideas (50%).
  + The winner team (having the highest compression ratio) will take the 100% of the competition grade. The team with least compression ratio will take 0% of the competition grade. The other teams will take a percentage of the compression grade depending on how far they are from the winner and loser teams.

# Deliverables

|  |  |
| --- | --- |
| Delivery | Due Date |
| Media | Sunday 21st May, 11:59pm. **To:** sarah.rashad.taha@gmail.com  **Email Subject:** [Multimedia][Semester] Team <your team no.> Example: [Multimedia][Semester] Team 12  **Attachment:** 1 zip file called: Team<your team no.>.zip Example: Team12.zip.  This attachment contains |
| Discussion | In the same week. Schedule will be announced later. |

**Note:**

Late media deliveries (after 11:59pm), wrong email subject or file names will make you lose marks. No modifications are allowed after the media delivery.

The 1 attachment zip file “**Team<your team no.>.zip**” must contain the following files:

* **Readme.txt** which contains the names of the team members.
* **Statistics.txt** which contains:
* For each text file, number of bytes in the compressed binary file and compression ratio of each.
* Total number of bytes in all compressed files and total compression ratio.
* **Report.pdf** of 2-5 pages which contains:
  + **[optional]** History of experiments you made before reaching the final algorithm.
* Clear description of your final compression and decoding algorithm. **Note:** The name of the algorithm is enough if you use a well-known algorithm as is but any modifications you made on the original algorithm or any parameters values that you use in the algorithm should be stated clearly.
* State clearly what is the content of each of the compressed files (if it contains only code words or any extra info) and what files sent to the decoding program are.