CSCI 599 – spring 2016 – Assignment 1 – Group 12

This report summarizes our work on first assignment. We have chosen

- Introduction:
- Getting Data:
- Task List:
    1. **Download and install Apache Tika**
       Done
    2. **Download and install D3.js**
       Done
    3. **Download the Amazon S3-based TREC-DD-Polar data**
       Done
    4. **Perform Byte Frequency Analysis on at 15 (14+ application/octet-stream) MIME types present in TREC-DD-Polar dataset**
       To be updated
    5. **Perform Byte Frequency Distribution Correlation and Byte Frequency Cross Correlation on the 15 (14+ application/octet-stream) MIME types present in the TREC-DD-Polar dataset**
       To be updated
    6. **Perform FHT analysis on your 15 (14+ application/octet-stream) MIME types present in the TREC-DD-Polar dataset**
       To be updated
    7. **Update Tika's MIME repository based on your BFA, BFC and FHT analyses**
       To be updated
    8. **Download and configure Tika Similarity and run it over your dataset**
       To be updated
    9. **Apply the ContentBased MIME detector in Tika to your TREC-DD-Polar data**
       To be updated