W205 Project timeline

W205 Project Timeline

- Week 11/01 - 11/07:
  - Decide the tweet information that we will be keeping for analysis and V1 analysis **- Sharmila**
  - Set up SparkSQL to update the 'tweets' table every time we get new data **- Kyle**
  - Creating the scraper of urls obtained from "spam" tweets **- Carlos**
  - Start: the check-up 2 pager **- Sharmila**
- Week 11/08 - 11/14
  - [Working on exercise 2] **- All**
  - Cron (or alternative) that starts flume, starts sparkSQL to update tweets table **- Carlos**
  - Storing the outputs of V1 analysis in S3 -> outputs must include a "spam"/"no spam" decision **- All**
  - **Publish the check-up 2 pager (to be submitted on Sunday Nov. 15th) - All**
- Week 11/15 - 11/21
  - [Finish exercise 2] **- All**
  - Start building rest API for retrieving data from S3 **- Sharmila/Carlos**
  - Start building dashboard for displaying data **- Kyle**
  - Start building browser plugin **- Kyle (Carlos wants to help!)**
- Week 11/22 - 11/28
  - Finish building rest API for retrieving data from S3 **- Sharmila/Carlos**
  - Finish building dashboard for displaying data **-Kyle**
  - Finish building browser plugin **- Kyle (Carlos wants to help!)**
  - End-to-end running/integration of project **- All**
- Week 11/29 - 12/05
  - Testing and optimization of the end-to-end code **- All**
  - Publish presentation **- All**
- **12/06 PROJECT DUE DATE**