

# 9-28-20 Status Report

## Leeds School of Business

Weekly status report for September 28-Oct 4

Scope	Schedule	Cost	Risks	Quality
-------	----------	------	-------	---------

## Key Performance Indicators

- **Schedule:** **ON SCHEDULE**
  - Schedule Variance: 0 Days
  - Percent complete : **15%**
- **Labor:** **ON BUDGET**
  - Labor hours: 9 hours/person
- **Administration:** **COMPLETE**
  - Team meetings, attendance, hour log

## Summary

We have developed a project charter and are awaiting approval from our sponsor next week. On top of this, we have created an API to efficiently handle the large dataset from Archive.org to start creating the Bloomberg TV corpus. This large dataset has gotten us to discuss storing the data in a DB for more efficient retrieval and storage methods.

## Work planned for this week

- Project charter
- Scrape and create BTV corpus

## Work completed this week

### Project Charter

We have filled out and created our project charter. Project Charter

This is subject to change and is awaiting approval from sponsor

### Scrape all BTV corpus

We created an API to efficiently download, store, and access information about segments provided by Archive.org. We can view metadata about segments, view segment transcripts as a whole or select minute by minute.

We have initial methods that organize text by speaker with varying success. We plan to implement better methods in the future.

### Started discussion of creating a DB

We have seen that there is a large amount of data we will have to deal with. Creating and storing it in a database will allow us to work much more efficiently. This link details the start of that discussion Database

## Plans for next week

- We have a lot of information to process. We think it will be better to store all this information in a DB for faster access. Continue conversation and decide which method we want to implement
  - see wiki docs on DB conversation
- Develop methods to identify repeated episodes in corpora

## Open Issues

- We have a large amount of data to process in the future. So much, that our own technology might not have the power to do so. We want to get set up with CU compute cluster so that our computing power will be much stronger.

## Deliverables and Milestones for sprint (9/22 to 10/6)

Deliverable or Milestone	WBS	Planned	Forecasted	Actual	Status
Project Charter	CAP-25	10/5/20	10/5/20	10/1/20	Completed
Basic BTV Corpus web scraping method	CAP-26	10/5/20	10/5/20	10/2/20	On Schedule

## Open Change Requests

Change Request Name	Change Request Number	Requested Date	Current Status