20-10-4 Status Report

Leeds School of Business

Weekly status report for Oct 4-Oct11

Scope	Schedule	Cost	Risks	Quality

Key Performance Indicators

- **Schedule:** `ON SCHEDULE`
  - Schedule Variance: 0 Days
  - Percent complete : **20%**
- **Labor:** `ON BUDGET`
  - Labor hours: 11 hours/person
- **Administration:** `COMPLETE`
  - Team meetings, attendance, hour log

Summary

To begin, our project charter was signed off by our project sponsor. Besides that, our TV scraper's output has been finalized. Now that we know how each tv show's output will look , we decided that a document database would serve our needs best. Therefore have implemented a mongo database to store our Bloomberg Tv corpora. Right now each person must have their own local version, but we are looking into alternatives to get it hosted online.

Work planned for this 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

Work completed this week

Project Charter Finalized

Professor Garcia agreed and signed off on our project charter

Database Design

We have decided to store our data in a mongoDB and have a working implementation

Some of us are new to this type of DB and need to do research on how to interact with them

SPIKE how to remove repeat shows in corpora

Based on the output for each show, we want to compare edit distance between current shows and ones ~2 day before it. If the edit distance is low enough the show will be thrown out

We have initial methods that seems like it could work but it is slow. We are looking into other potential methods such as cosine distance

Plans for next week

- Design document for possible interesting trends to analyze with the data
- Continue working on repeat episode
  - The edit distance seems to be too slow so we need to look into cosine distance

Open Issues

Deliverables and Milestones for sprint (10/6 to 10/20)

Deliverable or Milestone	WBS	Planned	Forecasted	Actual	Status
Model DB	CAP-30	10/13/20	10/13/20	10/10/20	Completed
Remove repeat shows in corpora	CAP-42	10/20/20	10/20/20		On Schedule
Sign Charter	CAP-50	10/13/20	10/13/20	10/7/20	Complete

Open Change Requests

Change Request Name	Change Request Number	Requested Date	Current Status