

10-11-20 Status Report

Leeds School of Business

Weekly status report for Oct 11-Oct 18

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

Key Performance Indicators

- **Schedule:** **ON SCHEDULE**
 - Schedule Variance: 0 Days
 - Percent complete : **25%**
- **Labor:** **ON BUDGET**
 - Labor hours: 12 hours/person
- **Administration:** **COMPLETE**
 - Sponsor meeting, TA Meeting, attendance, hour log

Summary

As we have implemented a functioning database, our goal for this week was to continue to clean the data and add more. We are working to clean the data by removing repeat episodes through cosine distance between transcripts. Also, we have decided to add more data by including FOX and CNBC archives. The more data we have, will hopefully lead to clearer insights when we start analyzing data.

Work planned for this 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

Work completed this week

Add FOX and CNBC to corpora

archive.org also contains FOX and CNBC TV archives. We added an option that allows the user to download the data from those two networks as well

Remove repeat shows in corpora

We have moved to using the cosine distance between vectors created from transcripts. If the cosine of the angle between them is high enough then the show will be thrown out

Plans for next week

- Implement a way to determine name entities in the transcripts
 - By doing this, we can better match what company/ who is speaking in the given interview
 - CAP-52
 - CAP-54

- Think of what kind of correlations we want to investigate
-

Open Issues

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

Deliverable or Milestone	WBS	Planned	Forecasted	Actual	Status
Add FOX and CNBC	CAP-46	10/20/20	10/20/20	10/15/20	Complete
Remove repeat shows in corpora	CAP-42	10/20/20	10/20/20	10/17/20	Complete

Open Change Requests

Change Request Name	Change Request Number	Requested Date	Current Status
