

10-18-2020 Leeds Status Report

Weekly status report for Oct 18-Oct 25

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

Key Performance Indicators

- **Schedule:** **ON SCHEDULE**
 - Schedule Variance: 0 Days
 - Percent complete : **30%**
- **Labor:** **ON BUDGET**
 - Labor hours: 12 hours/person
- **Administration:** **COMPLETE**
 - Sponsor meeting, attendance, hour log

Summary

This week was a little slower than the past weeks. We were successfully able to implement the duplicate detection and there is a design doc now discussing what the threshold will be. Other than that, most of the work was experimental, investigating good ways to match episodes with what company/subject they are about. This will hopefully be done using the wikipedia api, matching the name entities found in the episode to the company that is hopefully contained in the first few paragraphs of the wikipedia page.

Work planned for this 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

Work completed this week

Decide on cosine distance threshold

The given design doc gives a breakdown of all the comparisons that occur in the first 90 episodes. Two recommendations were given based on what information you want to detect in the duplicates

Investigate Wikipedia API

In the effort to refine matching tv episodes to what company/who is speaking we are investigating the wikipedia api which will hopefully match the name entities to what they belong to

This is in correlation to CAP-52 and will hopefully get us closer to an optimal solution

Plans for next week

- Find a remote service to host the database so everyone doesn't need their own local copy of the db
 - Continue CAP-52
-

Open Issues

- Project github needs a readme
 - Installation instructions text file can be removed
 - Dockerize our project
-

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

Deliverable or Milestone	WBS	Planned	Forecasted	Actual	Status
Interview matching process	CAP-52	11/3/20	11/3/20		In progress
	CAP-56				
Repeat show design doc for threshold	CAP-42	10/27/20	10/27/20		In progress

Open Change Requests

Change Request Name	Change Request Number	Requested Date	Current Status