Leeds Status Report Mar 29 to Apr 4

Weekly status report for Mar 29-Apr 4 2021

Scope	Schedule	Cost	Risks	Quality

Key Performance Indicators

- **Schedule:** **ON SCHEDULE**
  - Schedule Variance: 0 Days
  - Percent complete : **90%**
- **Labor:** **ON SCHEDULE**
  - Labor hours: 10 hours/person
- **Administration:** **COMPLETE**
  - Sponsor meeting, group meeting, attendance, hour log

Summary

We have started to hit some roadblocks that seem like there is no good solution for in the regex method of finding the company names on pages. Because of this, we have shifted focus to using our tensor flow model. This week we really decided to commit to the machine learning approach as we feel that this has strong potential to catch what we haven't been able to produce in our regex methods.

Work planned for this week

- Improve regex methods to find ~5-10% more company names per year - CAP-82
- Run Tensorflow on cluster - CAP-81

Work completed this week

Run Tensorflow on cluster - CAP-81

✓ Tensorflow model trained using this tutorial (https://tensorflow-object-detection-api-tutorial.readthedocs.io/en/latest/index.html). Everyone also completed 250 training images on their own of labeling where the company headers were. We were able to do this by using labelImg which allows us to create an xml mask saying where each header is on an image. Hopefully all of the training data we generated creates good results for us.

Improve regex methods to find ~5-10% more company names per year - CAP-82

⚠ We have ran into a lot of roadblocks. So this week we decided to take a step back from this and put a week into the tensorflow method. One idea that comes to mind to improve it though, is to switch from only taking all caps with co,corp, etc to removing all caps with a list of stop words. Although there are going to many stop words, this might work better since not all companies end in co,corp,etc.

Plans for next week

- Improve regex methods to find ~5-10% more company names per year - CAP-82
- Produce tensorflow results and examine- CAP-83

Open Issues

Deliverables and Milestones for sprint (3/21 to 4/4)

Deliverable or Milestone	WBS	Planned	Forecasted	Actual	Status
Produce tensorflow results and examine	CAP-83	4/11/21	4/11/21		In progress
Improve regex methods to find ~5-10% more company names per year	CAP-82	4/11/21	4/11/21		In Progress

Open Change Requests

Change Request Name	Change Request Number	Requested Date	Current Status