

Onboarding

1. Project goals

- The goal is to have you be an author on a published paper!
- Read our [previous paper](#)
- Previous paper [code and software](#)
- Read `docprop_replication/writing/project_description.pdf`

2. Communiation

- Slack for short informal messages (add to UMass NLP slack channel)
- Email for others
- In general, most messages and emails will be answered Mon-Friday 9-5
- Add them to NLP Google Groups for NLP reading group

3. Github

- Obtain access to SLANG lab github
- Get added as collaborator to `docprop_replication`
- `git clone git@github.com:slanglab/docprop_replication.git`

4. Google drive, keeping track of papers and progress

- [Google drive sharing link](#)
- ["MASTER docprop_replication"](#)

5. Hobbes

- email Dan Parker
- [SLANG lab internal documentation](#)
- tmux, ssh, scp

6. Managing tasks

- Github projects

7. Python environment

- Use Python 3,
- `virtualenv, virtualenvwrapper, pip` for installing packages

8. Research expectations

- 40 hours a week; flexible on working locations; being in the lab can be helpful when you're stuck;
- Research is really open-ended which can be hard to structure. If helpful, you a tool like Toggl to track

your hours and make sure you're working as much as you should.

- Attending talks, lab meetings, reading papers, reading background reading all counts as work! You should be working on your projects but also be expanding your knowledge of NLP and ML.
- Research is all about "narrowing the hypothesis space." Often times, you will have negative results and this is also important.
- Matt Gardner tutorial ["Writing Code for NLP Research"](#)

9. Yelp data locations

- training doc: `home/kkeith/docprop/yelp_data/train2000/train2000.json`
- test docs: `/home/kkeith/docprop/yelp_data/test`

10. Document, document, document

11. What to do in the next two days?

- Read previous paper and understand the methods (do some background reading with Kevin Murphy textbook!)
- Read Matt Gardner tutorial (above)
- Get python environment set up
- Preliminary task: yelp data on hobbess, use freq-e software; try to understand and replicate what is in the paper!

12. Meetings

- Katie is on campus 3-4 days a week
- Daily quick check-in the first couple of weeks (~2pm)
- Weekly meeting with Brendan set once he returns
- Weekly NLP reading group meeting
- Speak up when you don't know something! Not knowing is totally ok! Ask good questions! Helps everyone!