1st GP Meeting

October 12, 2018

Roadmap

- Course requirements
- MS Project options
  - Internship option
  - Thesis option
  - Project option
- Thesis planning

Courses

- Core courses:
  - Ling 550: Phonetics (AUT)
  - Ling 570: Shallow Processing for NLP (AUT)
  - Ling 571: Deep Processing for NLP (AUT)
  - Ling 566: Intro to Syntax (WIN)
  - Ling 572: Advanced Statistical Methods for NLP (WIN)
  - Ling 573: Systems/Applications (SPR)
- Elective courses: Linguistic, CompLing, CL or related
  - Each elective must have 3 or more credits
- Ling 600/700: Master's project

Course Requirements

- Must take (at least) 9 courses plus a capstone project:
  - If a required course is waived, it must be replaced with another course.
- 43 total credits required
  - If a required course is waived and replaced with a course with fewer credits, you may take add'l Ling 600 credits or another course to meet the 43-credit requirement.
- Minimum average 3.3 grade required in 3 groups
  - Ling 550, Ling 566, Ling elective
  - Ling 570, 571, 572
  - Ling 573, CompLing and CompLing or related electives
- Minimum grade: 2.7
  - Below 2.7 in required course: retake the course
  - Below 2.7 in elective: retake course or take alternative

MS Project Options

Internship Option

- Requirements: 10 credits Ling 600
  - Topic: relevant to CompLing or lang tech broadly
  - Workload: 200 hours, typically over a 4 month period

Internship Option

- Requirements: Documents
  - Pre-internship report (part 1): pre-search, 2-3 pages
    - Describe topics of interest, candidate companies, references
  - Pre-internship report (part 2): found job, 5 pages
    - Describe company, project, related literature
  - Evaluation letters (self and supervisor): end of internship
  - Post-internship report: typically 10-15 pages
    - Describe actual work, relation to CLMS courses, literature

Internship Option: Resources

- Advisor:
  - CLMS faculty advisor assigned based on interests
    - Advice on companies, contacts
    - Also will review all reports, evaluations
- Career services: resume, job search advice
- Internship lists, job posting database
  - Discussions, CLMS wiki, …

Internship Option: Deadlines

- 1/15: Pre-internship report (part 1)
- 1/31: Career Services visit completed
- mid-June: Internship starts
- 6/15: Pre-internship report (part 2)
- 7/20: Post-internship report (draft 1)
- 8/05: Self-evaluation (w/supervisor approval)
- 8/10: Supervisor evaluation form
- 8/10: Post-internship report (final)

Thesis Option

- Requirements: 10 credits Ling 700
  - Substantial novel research work
    - Comparable to workshop or conference paper
    - Typically system extension, implementation or evaluation
  - Initiative and independence:
    - Primary idea comes from and is executed by student
    - Must convince advisor and reader to take you on
  - Extended write-up: typically 30-50 pages

Thesis Structure

- Typical chapters:
  - Introduction
  - Literature review
  - Methodology
  - Algorithms, implementation
  - Experiments
  - Discussion
  - Conclusion and future work
- Follow University guidelines for electronic submission
  - Guidelines on web site

Thesis Option: Choices

- Choosing a topic:
  - Scope: Small enough to complete in a few months
    - Will almost always have to scale down from original idea
  - Novelty: Must contribute to knowledge of the field
    - New approach, new type of data (genre/language), new analysis

Thesis Option: Choices

- Choosing an advisor:
  - Must ‘sell’ advisor on your idea
  - Within advisor’s expertise, best if related to their work
  - Often CLMS faculty
    - But could be other faculty (EE, CSE) or in industry

Thesis Option: Deadlines (for summer graduation)

- late fall GP meeting: thesis discussion
- 1/3: Thesis proposal (first draft)
- 2/1: Thesis proposal (second draft)
- 3/1: Thesis proposal (final draft), agreed w/advisor
- 4/1: Literature survey
- 5/1: Methodology chapter
- 6/1: Interim results presentation to other students
- 7/1: Results chapter
- 7/7: First full draft
- 8/5: Second full draft
- 8/10: Final draft for signature and submission

Thesis Option: Notes

- Notes:
  - Deadlines are CLMS internal for summer graduation
    - Other timelines should be determined w/advisor
  - Must also meet grad school requirements
    - Degree application, electronic submission, etc
  - Reader: CLMS faculty or other
    - Either advisor or reader MUST be CLMS faculty

Thesis Option: Tips

- Tips:
  - A thesis ALWAYS takes longer than you think it will
  - Start early: register for thesis credits in SPR (or WIN)
  - “Double-dipping”: encouraged!
    - Use course projects to jumpstart, advance thesis work
    - Extend a project into a thesis
      - Talk to your advisor about scope

MS Project:
Project Option

Project Option

- Requirements: 10 credits of Ling 600
  - Topic: relevant to CompLing or lang tech broadly
  - Workload: 200 hours, typically over a 4-month period
  - Supervisor: Typically researcher at UW to oversee work

Project Option

- Requirements: Documents
  - Pre-project proposal: 5 pages
    - Plan for project, relation to CompLing, related literature
    - Must be approved by supervisor AND advisor
  - Evaluation letter: Self
  - Evaluation form: Supervisor: like internship
  - Post-project report:
    - Same as internship

Project Option: Deadlines

- 5/25: Pre-project proposal
- 6/15: Project proposal approved, supervisor agreed
- 7/20: Post-project report (first draft)
- 8/5: Self-evaluation (w/supervisor approval)
- 8/10: Supervisor's evaluation
- 8/10: Post-project report (final draft)

MS Project Comparison

	Thesis option	Internship option	Project option
Assigned an advisor?	No	Yes	Yes
Independent research?	Yes	No	No
Position in	university	industry	university
Normal length	6-9 months	3 months	3 months
Need to find	thesis topic advisor reader	Internship (supervisor)	Research Project (supervisor)
Main documents	thesis proposal thesis	pre-internship report post-internship report evaluation letters	pre-project report post-project report evaluation letters
Followup	publications	full-time job	(publications)

Choosing an Option

- Typically:
  - Internship: 2/3; Thesis: 1/5; Project: remainder
- Internship:
  - Good for work experience, networking
  - Can lead to permanent position
  - Good choice if want an industry job in the short-term
- Thesis:
  - More work, more initiative, tighter deadlines
  - Good choice if plan to apply to a Ph.D. program
- Project:
  - Contribute to on-going work
  - Good option for those with full-time non-NLP jobs

First Steps: Groups for Graduation

- Group 1:
  - Planning to start GP process this year: 2018-2019
  - Attend all upcoming GP meetings
- Group 2:
  - Went through GP process already
- Group 3:
  - Plan to go through GP process in later year
- Reply to corresponding thread on Canvas discussions:
  - CLMS Graduate Planning

Final Notes

- Talk to your advisor about your plans
  - Internship, thesis, or project
  - We're here to help!!
- Pay attention to the deadlines
  - Key to graduating on time
  - Some flexibility, but must be approved by advisor

Three dimensions of a Thesis

- Key components of a thesis
  - Dimension 1: Task
  - Dimension 2: Algorithms
  - Dimension 3: Data/Domain/Language

Dimension 1: Task

- **What problem** are you trying to solve?
- Possible tasks: (some topics from Ling 570, 571)
  - Word segmentation
  - Morphological analysis
  - POS tagging
  - Named entity recognition
  - Grammar induction, etc….
- Application: (some covered in Ling 573)
  - Q&A, MT, ASR, sentiment detection, topic detection, etc

Dimension 2: Algorithms

- **How** are you going to solve the problem?
- Classes of approach:
  - Symbolic approaches (aka rule-based approaches)
  - Corpus-based approaches (machine learning, Ling 572)
    - Supervised learning:
      - Decision trees,, MaxEnt, SVM, neural nets
    - Semi- or Un-supervised approaches:
      - Co-training, EM, Clustering, etc
  - Hybrid approaches:
    - Mix symbolic and corpus-based methods

Dimension 3: Data/Domain

- **What** materials will you use to evaluate your method?
  - **What data and evaluation techniques** will you use?
- Many types of data:
  - Raw data, treebanks, FrameNet, tagged corpora
  - Monolingual, bilingual, multi-lingual
  - Newswire, conversational speech
    - Limited domain: travel, bio-medical

Projects

- Draw from each of the dimensions
- What task?
- What algorithm?
- What data/domain/language?
- Novelty:
  - Novel aspects can come in any dimension or by novel combination, e.g. of data and algorithm

Research Interests

- See Appendix C in handout