Doing Resarch in Natural Language Processing

Session 1: Course Overview

Frank Keller and Hao Tang

20 September 2023

School of Informatics
University of Edinburgh
keller@inf.ed.ac.uk

Overview

Icebreaker

We will start with a quick icebreaker task:

- Please briefly introduce yourself: name, pronouns, where you're from, which languages you speak.
- Say something about your academic background and research interests.
- Very briefly talk about a research project you've done recently (course project, MSc project, internship, etc.).

Introduction

Introduction

- This is a course on research methods.
- It covers generic things such as scientific writing, presentation skills, peer review.
- but also methods specific to NLP and related fields:
  - replication
  - cluster computing
  - online experiments
  - ethical issues
- critically reading the scientific literature is another important skill; we will develop this in reading group sessions.
- We will also talk about public engagement and about working with industry.

Relation to other Courses

- This course is designed so that you can apply the skills you learn in the Group Project and in the Individual Project.
- In addition, this course includes its own small project, the *replication project* in semester 1.
- Aspects of the Individual Project (proposal writing, progress presentation) are integrated into this course in semester 2.

Course Mechanics

Course Format

- This is not a lecture course.
- But there will be tutorial-style material, some presented by the lecturers, some by guest speakers.
- We will rely on student participation for many aspects of the course.
- This can take the form of classroom discussion, contribution to Piazza, or student presentations.
- On some aspects of the course, students will work together in small groups (2–3 students).

Course Delivery

- Two sessions per week. These will happen in person, but will also be recorded and live streamed.
- Lecturers: Frank Keller and Hao Tang.
- Teaching assistant: Gautier Dagan (cohort 2021).
- The TA will hold weekly office hours; details tba.

Course Materials

Like most Informatics courses, we will use a combination of Informatics OpenCourse, Learn, and Piazza:

- https://opencourse.inf.ed.ac.uk/nlp-dr contains course materials (slides, readings, coursework exercises).
- https://www.learn.ed.ac.uk/ultra/courses/_108488_1/outline contains "official" stuff like coursework deadlines, submission portal, lecture recordings.
- Piazza is linked from Learn will be used for discussion of course materials and logistics.

Assessment

Assessment

This course is coursework-only, there is no exam. There will be three pieces of assessed coursework:

1. A research report on a replication study (semester 1), worth 40%.
2. A proposal for your individual project (semester 2), worth 30%.
3. A presentation reporting on the progress of your IP (semester 2), worth 30%.

Please see the Learn course page for submission deadlines.

Replication Study

- As part of this course, we will critically discuss the issue of *replicability* of research results.
- As a practical exercise, you will replicate a result from the literature.
- You will write up the result as an (assessed) report, and present your results to the class.
- This exercise will be done in pairs. Every pair will replicate a different result.

Replication Study

Timeline of the replication study:

- Week 4: We release a list of replication papers.
- Week 4: Students submit a ranked list of three paper they want to replicate.
- Week 5: We assign papers to students and form groups. We issue instructions for the assignment.
- Week 9: Students submit their replication reports (deadline: 13 Nov).

Your Immediate Tasks

Reading

For the writing skills part, we will use Alley's (2018) *The Craft of Scientific Writing*.

Alley covers scientific writing generally, but is also very relevant to NLP. The e-book is linked from the course Learn site.

We will set reading for each session; a reading list will appear on Learn.

There will also be small writing exercises. These will typically be announced on Learn or on Piazza.

Research Papers

- We will have a number of *reading group* sessions.
- In these, 2–3 students lead a discussion of a research paper.
- We would like to discuss papers of broad relevance to language and computation.
- You should start thinking about which papers you'd like to present.
- The TA will soon start collecting your paper suggestions.
- The lecturers will vet them :)

We will look at the content of these paper, but also at their *writing* (structure, style, abstract, intro, figures/tables) and at their *replicability.*

References

Alley, Michael. 2018. *The Craft of Scientific Writing*. Springer, New York, NY, 4 edition.