



Annotation Projects

Jan 30, 2018



+ Major phases of the Annotation Project

2

- Phase 1: Defining the task and model (thru 2/9)
 - Group contract
 - Annotation goal
 - Task description and corpus selection
 - Draft annotation specification
- Phase 2: Test and refine and the specification (thru 2/27)
 - Full Annotation Schema
 - Full Task specification
 - Guidelines
- Phase 3: Manage annotators (Thru 3/27)
 - Weekly progress reports
 - Annotation report
 - Note: We'll decide at the time whether annotation will continue over break or you start the gold standard process over break (or both ;-)
- Phase 4: Gold standard and Learning model (thru ~5/4)
 - Gold standard release
 - Presentation
 - Final paper

+ Phase 1

- Defining the task and model (thru 2/9)
 - Select the corpus
 - Determine where the data will be kept and how you plan to access it
 - Data exploration: How well does it fit your goals and model?
 - Put the data in the correct format for MAE
- Deliverables
 - Group contract
 - Annotation goal
 - Task description and corpus selection
 - Draft annotation specification

+ Phase 2

- Test and refine and the specification (thru 2/27)
 - Write the first draft of the specification and DTD
 - Select a set of “documents” for everyone to annotate & compare results
 - Assign a larger set to find canonical and edge cases for the guidelines
 - Adjust model, specifications, and guidelines
 - Figure out how long it will take to annotate each document
- Deliverables
 - Full Annotation Schema
 - Full Task specification
 - Guidelines

+ Phase 3

- Manage annotation (Thru 3/17)
 - Create a data flow pipeline so your annotators know how they get the data, where they put it, exactly what's assigned to them
 - Train annotators, set up a communication channel for questions
 - Monitor annotators, answer questions in a timely fashion
 - Compute IAA, adjust model, specification and guidelines
 - Manage updates to ensure annotators are using the most recent version of the specification
- Deliverables
 - Weekly progress reports
 - Annotation report

+ Phase 4

- Gold standard and Learning model (thru ~5/4)
 - Determine how to most efficiently arrive at a single set of labeled data (full match, voting, when to view by hand)
 - Create “Gold standard”
 - Determine what classification or other learning technique that best fits your task goal, form of annotation, and amount of data
 - Create a model and run a test. Report results in presentation and final paper
- Deliverables
 - Gold standard release
 - Presentation
 - Final paper