

Course Project

Complete document will be posted on website this week

- Build a sequence tagger,
for Twitter named entity recognition
 - We provide training and test data
 - Online submission of predictions and leaderboard
- Work by yourself, or in group of 2
- Deliverables
 - 11/16, milestone:
rudimentary system, initial results, successful online submission, short report
 - 12/5, final report, model and code

Course Project Requirements

- 1. System building
 - We will provide starter code to use open-source tagging software, CRFSuite, which accepts features as input.
 - You implement a core system:
lexical, affix, n-gram features
 - Extensions (must do one if group of two; more are extra credit):
 - Name list features
 - Unsupervised learning features
 - POS or other NLP system input features
 - Different sequence tagging algorithms
 - EXTRA CREDIT: especially good tagger performance
- 2. Exploration
 - Analyze what your model has learned
 - Ablation tests: what features are important?
 - EXTRA CREDIT:
Run your tagger to analyze a large corpus of tweets that we provide.
Analyze something you find interesting (e.g. most popular names over time)

Rest of semester

- Only 5 problem sets total. PS3 due next Tuesday. (A short Part 2 is being posted this evening)
- PS2 grading is almost done -- sorry for the delay
- Project
- Final exam on 12/11