Tasks, somewhat prioritized

* Evaluation
  + Incorporate XValidation machinery and look at the training sets instead of trial sets for both domains
  + Try other classifiers: SVM, RF, NB
* add more features
  + prev & next word weren’t actually added to dictionary
  + expand n-gram window
  + add the more complicated sentiment dictionary
  + all caps
  + initial cap
  + hashtag (not for aspect task)
* start on the sentiment task over spring break
  + incorporate handling of negation
  + feed sentiment classification back into aspect term extraction – if there is only neutral text then there is no aspect term
  + perhaps divide into two: subjective vs objective then pos/neg/mixed
* (C) figure out how to best generate figures and tables for inclusion in a paper
* does sklearn have any sequence classification?

Some ideas

* could also use a NP chunker as additional evidence for aspect term extraction
* really do sequence modeling – seems that the head is stronger evidence than earlier words in the phrase
* are verbs ever extracted? – could print out if we find one
* add the actual sentiment indicator words (those from the lexicon) as features for aspect term id, but explore different ways to do it
  + closest
  + all
  + if in same (hi-level) NP
  + the actual word, its polarity, or just the presence of such a word
* use the classifier from the other domain as additional evidence (for all the tasks but starting with aspect term extraction)
* a feature that indicates we’ve seen the word but it’s never been in an aspect term (or the proportion of the time it’s been in an aspect term)
* sentiment for the aspect terms
  + as part of it, should really figure out which sentiment-laden words (like “sluggish” modify which aspect terms (like driver, in the phrase sluggish driver)
  + types of patterns
    - the X was Y (steak was delicious)
    - Y X (poor service)
  + How to use sentences for which there was no aspect term & thus no sentiment? Words used in a neutral setting, perhaps?
  + Sentiment classification (& confidence/strength) for overall sentence
  + PMI between sentiment words and heads of aspect phrases?
  + Dependency parse closest words
* What helps in WSD?
  + Character ngrams?

POS tagging notes:

* The create\_exs method now assumes each aspect token appears only once in a given sentence, rather than worrying about which occurrence of the token actually is the aspect term if there are multiple ones. See next bullet for problems faced with dealing with both POS tagging and punctuation in finding the aspect terms (doable but messy!)
* the first version of the code counts (create\_one\_withPOS) the character indexes incorrectly when punctuation is involved (for the POS tag case, may also be true for non-pos case, not sure)
  + The “to” position in the xml is the position after the EOW, with punctuation stripped, if any
  + Could use a dictionary with start index as keys, (word, BIO-tag) as values, and include punctuation in that? Then iterate through range(len(sentence) and see if those are in the dict keys. But how to get that dictionary
    - Assume punctuation always attached to a word? So don’t add 1 in that case, just add len of punct? Or see if the punctuation is in the split sentence at the appropriate position.
    - Or iterate through the original characters. Can tokenize give us access to the original positions?

Sentiment lexicon notes:

* to use is Bing Liu’s lexicon and Wilson etal’s lexicon
* B.Liu’s is just word list for pos & neg ```````````
* Wilson etals associates words with pos tags and stemmed versions of the words, so to use that well need stemming
* Not sure yet how to use s.lex. entries for evidence in aspect term extraction, but presence within a window is a first reasonable cut, and of course it should be used for the polarity tagging task

**To get started in interpreter:**

sys.path.extend(['/users/cindi/semeval'])

import semevalTask4

**Try train and test on the trial data first to see how it works.**

ChunkParse score on restaurants-trial 80/20 split with no POS tagging (~2/23)

IOB Accuracy: 89.0%

Precision: 75.0%

Recall: 24.0%

F-Measure: 36.4%

Added POS tags, 3/1: Slightly hurts precision but awesome (in comparison to prev) on recall – makes sense, can incorporate syntax info

IOB Accuracy: 91.7%

Precision: 73.3%

Recall: 44.0%

F-Measure: 55.0%

Adding sentiment dictionary: helps precision quite a bit while hurting recall and F-measure:

IOB Accuracy: 91.4%

Precision: 83.3%

Recall: 40.0%

F-Measure: 54.1%

**Results for subtask 2**

* 3/9, initial classifier = maxent iis, features as below, accuracy = 64%!
  + Overall sentence positive & negative sentiment score
  + Number of aspect terms
  + POS, word, iob, sentiment of word preceding & following aspect phrase
  + Word, pos & iob of head of aspect phrase