HW09: Sequence Labeling and Bootstrapping

Hand in at: `http://www.cs.utah.edu/~hal/handin.pl?course=cmsc723`. Remember that only PDF submissions are accepted. We encourage using LaTeX to produce your writeups. See `hw00.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running "`pdflatex hw00.tex`".

1. In class we discussed using the structured perceptron to do named entity recognition. Sketch an algorithm for doing structured perceptron for *parsing.* What features would you use, and what algorithm(s) would you use for solving "argmax?"

2. Why might it be difficult to scale the DIRT algorithm to work at the scale of the web (i.e., billions of documents)? In particular, are there aspects that scale *super-linearly* in the data set size?