Today's Topic

• Reminder Homework 4 due tomorrow midnight!
• Install tregex
Install tregex

https://nlp.stanford.edu/software/tregex.shtml
Penn Treebank documentation

1. Tagging Guide
2. Arpa94 paper
3. Parse Guide
tregex

- Select the PTB directory
  - TREEBANK_3/parsed/mrg/wsj/
- Browse

Deselect any unwanted files
tregex

• Search
  • NP-SBJ << (dominates) vs. < (immediately dominates) NNP

Mr. Vinken, 61 years old, will join the board as a noneexecutive director Nov. 29.

Mr. Vinken is chairman of Elsevier N.V., the Dutch publishing group.
tregex

Tregex Pattern Syntax and Uses

Using a Tregex pattern, you can find only those trees that match the pattern you're looking for. The following table shows the symbols that are allowed in the pattern, and below there is more information about using these patterns.

Table of Symbols and Meanings:
A << B
  A dominates B
A >> B
  A is dominated by B
A < B
  A immediately dominates B
A > B
  A is immediately dominated by B
A = B
  A is a sister of B (and not equal to B)
A , B
  A precedes B
A . B
  A immediately precedes B
A , B
  A follows B
A . B
  A immediately follows B
A < , B
  B is a leftmost descendent of A
A <= B
  B is a rightmost descendend of A
A >> , B
  A is a leftmost descendent of B
A > , B
  A is a rightmost descendend of B
A < , B
  A is the first child of A
A > , B
  B is the first child of A
A = , B
  A is the first child of B
A <= B
  B is the last child of A
tregex

- The best introduction to Tregex is the brief powerpoint tutorial for Tregex by Galen Andrew.
  - The_Wonderful_World_of_Tregex.ppt
tregex

• useful command line tool:
  • diff <file1> <file2>