LING 581: Advanced Computational Linguistics

Lecture Notes
March 20th
Today's Topics

• EVALB Graph
• WordNet
Homework 7 Review

- Thanks for the PTB WSJ EVALB graphs so far...
- Don't forget to run data points and submit yours as Homework 7.

22 minutes for Section 23 (2241 sentences)
A HP workstation
Two Intel Xeon CPUs
E5-2687W @ 3.10GHz
16 cores used
split –l 25 used
using GNU parallel
Homework 7 Review

```bash
#!/bin/bash
# assumes input files to parser are in scratch
# assumes $1 is comma-separated list of suffixes without braces
# assumes $2 specifies the obj.gz file
# assumes $3 = the \$scratch subdirectory
# assumes subdirectories
# ptb (PTB parsing/training data)
# dbp (Bikel Collins parser)

10SECONDS=0

for x in `eval echo \"$1\"`
  do
date
  dbp/bin/parse 1500 dbp/settings/collins.properties ptb/\$x.obj.gz \$3/40wsj_23-$x
  done

for ((i=2; i<=$3SECS; i=i+1))
  do
date
  echo \"rm -f $SCRATCH/40wsj_23-*.parsed\"
  cat $SCRATCH/40wsj_23-*.parsed > 40wsj_23-$i.parsed
  done
```

run-chunk.sh
Homework 7 Review

- split –l 100
- split –l 50
- split –l 25
Homework 7 Review

- `split -l 50 job size`

- `split -l 25 job size`

![Graph 1: Job size for PTB WSJ Section 23 partitioned by 50](image1)

![Graph 2: Job size for PTB WSJ Section 23 partitioned by 25](image2)
Homework 7 Review

• my EVALB results ..
Today’s Topic

• WordNet 3.0
  • (3.1 the latest version but only online or the database files only)
  • http://wordnetweb.princeton.edu/perl/webwn
  • **Homework**: do the installations and use the programs in this lecture...
WordNet 3.0

• [https://wordnet.princeton.edu](https://wordnet.princeton.edu)
NLTK and WordNet

http://www.nltk.org/howto/wordnet.html

```python
>>> from nltk.corpus import wordnet as wn
>>> wn.synsets('cat')
[Synset('cat.n.01'),
 Synset('guy.n.01'),
 Synset('kat.n.01'),
 Synset('cat-o-nine-tails.n.01'),
 Synset('caterpillar.n.02'),
 Synset('big_cat.n.01'),
 Synset('computerized_tomography.n.01'),
 Synset('cat.v.01')] >>> s = wn.synsets('cat') >>> s[6] Synset('big_cat.n.01')

>>> s[6].lemma_names()
['big_cat', 'cat']

>>> s[6].lemma_names('fra')
['chat', 'fauve', 'félin']

>>> s[6].hypernyms()
[Synset('feline.n.01')]

>>> s[6].hypernyms()[0].hypernyms()
[Synset('carnivore.n.01')]

>>> s[6].hypernyms()[0].hypernyms()[0].hypernyms()
[Synset('placental.n.01')]
```
WordNet 3.0

• Download: https://wordnet.princeton.edu/download/current-version

WordNet 2.1 for Windows
WordNet browser, command-line tool, and database files with InstallShield self-extracting installer:
Download: WordNet-2.1.exe

WordNet 3.0 for UNIX-like systems (including: Linux, Mac OS X, Solaris)

Prolog version of WordNet 3.0
ANSI Prolog version of the WordNet database.
Download: WNprolog-3.0.tar.gz
WordNet 3.0

• Steps
  • Download.. from website

  • README
    • Beginning with Version 2.1, we changed the Unix package to a GNU Autotools package.
    • The WordNet browser makes use of the open source Tcl and Tk packages. Tcl and Tk must be installed BEFORE you compile WordNet.
    • You must also have a C compiler before installing Tcl/Tk or WordNet. WordNet has been built and tested with the GNU gcc compiler.
WordNet 3.0

• Steps
  • GNU Autoconf
  • configure
    cd Downloads/WordNet-3.0
    ./configure
    • checking for gcc... gcc
    • config.status: creating Makefile
    • checking for Tcl configuration... configure: WARNING: Can't find Tcl configuration definitions
  • INSTALL
    If you're running OS X and installed the Aqua Tcl/Tk package from the web site above, use the following settings:
    --with-tcl=/System/Library/Frameworks/Tcl.framework
    --with-tk=/System/Library/Frameworks/Tk.framework

  need to have Xcode installed on Mac

  OSX 10.9

  make /usr/bin/make
WordNet 3.0 on MacOS
WordNet 3.0

- GNU Autoconf

./configure -with-
tcl=/Applications/Xcode.app/Contents/Developer/Platforms/MacOSX.platform/Developer/SDKs/MacOSX10.9.sdk/System/Library/Frameworks/Tcl.framework -with-
tk=/Applications/Xcode.app/Contents/Developer/Platforms/MacOSX.platform/Developer/SDKs/MacOSX10.9.sdk/System/Library/Frameworks/Tk.framework/

cHECKING FOR GCC... gcc

... checking for Tcl configuration... found
/Applications/Xcode.app/Contents/Developer/Platforms/MacOSX.platform/Developer/SDKs/MacOSX10.9.sdk/System/Library/Frameworks/Tcl.framework/tclConfig.sh

... checking for Tk configuration... found
/Applications/Xcode.app/Contents/Developer/Platforms/MacOSX.platform/Developer/SDKs/MacOSX10.9.sdk/System/Library/Frameworks/Tk.framework/tkConfig.sh

... checking for existence of
/Applications/Xcode.app/Contents/Developer/Platforms/MacOSX.platform/Developer/SDKs/MacOSX10.9.sdk/System/Library/Frameworks/Tcl.framework/tclConfig.sh... loading

... checking for existence of
/Applications/Xcode.app/Contents/Developer/Platforms/MacOSX.platform/Developer/SDKs/MacOSX10.9.sdk/System/Library/Frameworks/Tk.framework/tkConfig.sh... loading

configure: creating .config.status

config.status: creating Makefile
WordNet 3.0

WordNet is now configured

Installation directory: /usr/local/WordNet-3.0

To build and install WordNet:

make
make install

To run, environment variables should be set as follows:

PATH - include ${exec_prefix}/bin
WNHOME - if not using default installation location, set to /usr/local/WordNet-3.0

See INSTALL file for details and additional environment variables which may need to be set on your system.
WordNet 3.0

• make

```bash
In file included from tkAppInit.c:16:
/Applications/Xcode.app/Contents/Developer/Platforms/MacOSX.platform/Developer/SDKs/MacOSX10.9.sdk/usr/include/tk.h:78:11: fatal error: 'X11/Xlib.h' file not found
#  include <X11/Xlib.h>

1 error generated.
```

recall: from your EVALB compilation a .h file means a C compiler header file
X11 is now under /opt/X11,
include files are in /opt/X11/include
WordNet 3.0

• configure
  ./configure --with-
tcl=/System/Library/Frameworks/Tcl.framework/ --with-
tk=/System/Library/Frameworks/Tk.framework/ CFLAGS=--
I/opt/X11/include

• make
  gcc -I/opt/X11/include -o wishwn wishwn-tkAppInit.o
  wishwn-stubs.o -L../lib -lWN -F/System/Library/Frameworks -
  framework Tk -F/System/Library/Frameworks -framework Tcl -
  lpthread -framework CoreFoundation -framework Cocoa -
  framework Carbon -framework IOKit -lpthread -framework
  CoreFoundation

• See also
  • https://stackoverflow.com/questions/23658393/installing-wordnet-on-mac-10-9-2
WordNet 3.0

• **sudo make install**

  ```
  ...  
  Making install in src
  test -z "/usr/local/WordNet-3.0/bin" ||
  /Users/sandiway/Downloads/WordNet-3.0/install-sh -d
  
  
  /usr/local/WordNet-3.0/bin/
  /usr/bin/install -c 'wn' 
  /usr/bin/install -c 'wishwn' 
  /usr/bin/install -c 'wnb' 
  ```

  • Running the WordNet browser:

  ```
  /usr/local/WordNet-3.0/bin/wnb
  /usr/local/WordNet-3.0/bin/wnb: line 3: wishwn: command not found
  ```
WordNet 3.0

- PATH to wnb (WordNet browser)
  ```
  export PATH=/usr/local/WordNet-3.0/bin:$PATH
  which wnb
  ```
- `/usr/local/WordNet-3.0/bin/wnb`

wnb

![WordNet 3.0 Browser](image)

requires platform packages that have:
- gcc
- make
- configure
- TCL/TK

The noun hello has 1 sense (first 1 from tagged texts)

1. (1) **hello**, hullo, hi, howdy, how-do-you-do -- (an expression of greeting; "every morning they exchanged polite hellos")
WordNet 3.1 vs 3.0
Perl Interface to WordNet 3.0

• http://www.cpan.org/

look for WordNet QueryData
Perl Interface to WordNet 3.0

http://search.cpan.org/~jrennie/WordNet-QueryData-1.49/QueryData.pm
Perl Interface to WordNet 3.0

• Use the cpanm command or manually (see next slides)

dhcp-10-142-168-95:5.18 sandiway$ sudo cpanm WordNet::QueryData
--> Working on WordNet::QueryData
Fetching http://www.cpan.org/authors/id/J/JR/JRENNIE/WordNet-QueryData-1.49.tar.gz ... OK
Configuring WordNet-QueryData-1.49 ... OK
Building and testing WordNet-QueryData-1.49 ... OK
Successfully installed WordNet-QueryData-1.49
1 distribution installed
Perl Interface to WordNet 3.0
Perl Interface to WordNet 3.0

```
cd WordNet-QueryData-1.49
perl Makefile.PL
  • Checking if your kit is complete...
  • Looks good
  • Writing Makefile for WordNet::QueryData
make
• cp QueryData.pm blib/lib/WordNet/QueryData.pm
• Manifying blib/man3/WordNet::QueryData.3pm

make test
PERL_DL_NONLAZY=1 /usr/bin/perl "-Iblib/lib" "-Iblib/arch" test.pl
ok 1
ok 2
ok 3
ok 4
ok 5
...
ok 104
ok 105
ok 106

sudo make install
Installing /Library/Perl/5.18/WordNet/QueryData.pm
Installing /usr/local/share/man/man3/WordNet::QueryData.3pm
Appending installation info to /Library/Perl/Updates/5.16.2/darwin-thread-multi-2level/perllocal.pod
```
Perl Interface to WordNet 3.0

NAME

WordNet::QueryData - direct perl interface to WordNet database

SYNOPSIS

```perl
use WordNet::QueryData;
my $wn = WordNet::QueryData->new( noload => 1);

print "Synset: ", join("", ", $wn->querySense("cat#n#7", "syns")), "\n";
print "Hyponyms: ", join("", ", $wn->querySense("cat#n#1", "hypo")), "\n";
print "Parts of Speech: ", join("", ", $wn->querySense("run")), "\n";
print "Senses: ", join("", ", $wn->querySense("run#v")), "\n";
print "Forms: ", join("", ", $wn->validForms("lay down#v")), "\n";
print "Noun count: ", scalar($wn->listAllWords("noun")), "\n";
print "Antonyms: ", join("", ", $wn->queryWord("dark#n#1", "ants")), "\n";
```
Perl Interface to WordNet 3.0

perl wnqdata.perl
print "Synset: ", join(" ", $wn->querySense("cat#n#7", "syns"))), "\n";
• Synset: big_cat#n#1, cat#n#7
print "Hyponyms: ", join(" ", $wn->querySense("cat#n#1", "hypo"))), "\n";
• Hyponyms: domestic_cat#n#1, wildcat#n#3
print "Parts of Speech: ", join(" ", $wn->querySense("run"))), "\n";
• Parts of Speech: run#n, run#v
print "Senses: ", join(" ", $wn->querySense("run#v"))), "\n";Senses:
• run#v#1, run#v#2, run#v#3, run#v#4, run#v#5, run#v#6, run#v#7, run#v#8, run#v#9, run#v#10, run#v#11, run#v#12, run#v#13, run#v#14, run#v#15, run#v#16, run#v#17, run#v#18, run#v#19, run#v#20, run#v#21, run#v#22, run#v#23, run#v#24, run#v#25, run#v#26, run#v#27, run#v#28, run#v#29, run#v#30, run#v#31, run#v#32, run#v#33, run#v#34, run#v#35, run#v#36, run#v#37, run#v#38, run#v#39, run#v#40, run#v#41
Perl Interface to WordNet 3.0

perl wmqdata.perl
print "Forms: ", join("", ", ", $wn->validForms("lay down#v")), "\n";
• Forms: lay_down#v, lie_down#v
print "Noun count: ", scalar($wn->listAllWords("noun")), "\n";
• Noun count: 117798
print "Antonyms: ", join("", ", ", $wn->queryWord("dark#n#1", "ants")), "\n";
• Antonyms: light#n#9
wnb

- WordNet browser: code written in tcl/tk, runs under wish
wnb

• Setup:
  • export PATH=/usr/local/WordNet-3.0/bin:$PATH
  • wnb
  • Options: see sense numbers
    (we’ll need these later)
• Hypernyms (going up the hierarchy):

<table>
<thead>
<tr>
<th>Sense 1</th>
<th>cat#1, true cat#1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>⇒ feline#1, felid#1</td>
</tr>
<tr>
<td></td>
<td>⇒ carnivore#1</td>
</tr>
<tr>
<td></td>
<td>⇒ placental#1, placental mammal#1, eutherian#1, eutherian mammal#1</td>
</tr>
<tr>
<td></td>
<td>⇒ mammal#1, mammalian#1</td>
</tr>
<tr>
<td></td>
<td>⇒ vertebrate#1, craniate#1</td>
</tr>
<tr>
<td></td>
<td>⇒ chordate#1</td>
</tr>
<tr>
<td></td>
<td>⇒ animal#1, animate being#1, beast#1, brute#2, creature#1, fauna#2</td>
</tr>
<tr>
<td></td>
<td>⇒ organism#1, being#2</td>
</tr>
<tr>
<td></td>
<td>⇒ living thing#1, animate thing#1</td>
</tr>
<tr>
<td></td>
<td>⇒ whole#2, unit#6</td>
</tr>
<tr>
<td></td>
<td>⇒ object#1, physical object#1</td>
</tr>
<tr>
<td></td>
<td>⇒ physical entity#1</td>
</tr>
<tr>
<td></td>
<td>⇒ entity#1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sense 2</th>
<th>guy#1, cat#2, hombre#1, bozo#2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>⇒ man#1, adult male#1</td>
</tr>
<tr>
<td></td>
<td>⇒ male#2, male person#1</td>
</tr>
<tr>
<td></td>
<td>⇒ person#1, individual#1, someone#1, somebody#1, mortal#1, soul#2</td>
</tr>
<tr>
<td></td>
<td>⇒ organism#1, being#2</td>
</tr>
</tbody>
</table>
wnb

• Hyponyms (going down the hierarchy):

3 of 8 senses of cat

Sense 1
cat#1, true cat#1
  => domestic cat#1, house cat#1, Felis domesticus#1, Felis catus#1
  => wildcat#3

Sense 2
guy#1, cat#2, hombre#1, bozo#2
  => sod#4

Sense 7
big cat#1, cat#7
  => leopard#2, Panthera pardus#1
  => snow leopard#1, ounce#3, Panthera uncia#1
  => jaguar#1, panther#1, Panthera onca#1, Felis onca#1
  => lion#1, king of beasts#1, Panthera leo#1
  => tiger#2, Panthera tigris#1
  => liger#1
  => tiglon#1, tigon#1
  => cheetah#1, chetah#1, Acinonyx jubatus#1
  => saber-toothed tiger#1, sabertooth#1
wnb

• Holonyms (part-of hierarchy):

1 of 8 senses of cat

Sense 7
big cat#1, cat#7

MEMBER OF: Felidae#1, family Felidae#1
MEMBER OF: Carnivora#1, order Carnivora#1
MEMBER OF: Eutheria#1, subclass Eutheria#1
MEMBER OF: Mammalia#1, class Mammalia#1
MEMBER OF: Vertebrata#1, subphylum Vertebrata#1, Craniata#1, subphylum Craniata#1
MEMBER OF: Chordata#1, phylum Chordata#1
MEMBER OF: Animalia#1, kingdom Animalia#1, animal kingdom#1
wnb

- Meronyms (has parts, plus inherited):

```
Sense 1
  cat#1, true cat#1
    => feline#1, felid#1
      HAS PART: paw#1
      HAS PART: pad#7
    => carnivore#1
      => placental#1, placental mammal#1, eutherian#1, eutherian mammal#1
      => mammal#1, mammalian#1
      HAS PART: coat#3, pelage#1
      HAS PART: hair#4, pilus#1
    => vertebrate#1, craniate#1
      HAS PART: belly#5
      HAS PART: tail#1
      HAS PART: dock#6
      HAS PART: caudal appendage#1
      HAS PART: rib#2, costa#2
      HAS PART: costal cartilage#1
      HAS PART: thorax#2, chest#1, pectus#1
      HAS PART: sternum#1, breastbone#1
      HAS PART: gladiolus#2, corpus sternum#1
      HAS PART: manubrium#1
      HAS PART: xiphoid process#1
```
• Meronyms (has parts, plus inherited):

5 senses of car

Sense 1
car#1, auto#1, automobile#1, machine#6, motorcar#1
HAS PART: accelerator#1, accelerator pedal#1, gas pedal#1, gas#5, throttle#2, gun#6
HAS PART: air bag#1
HAS PART: auto accessory#1
HAS PART: automobile engine#1
HAS PART: exhaust#2, exhaust system#1
HAS PART: exhaust manifold#1
HAS PART: exhaust pipe#1
HAS PART: exhaust valve#1
HAS PART: silencer#2, muffler#1
HAS PART: tailpipe#1
HAS PART: automobile horn#1, car horn#1, motor horn#1, horn#11, hooter#2
HAS PART: horn button#1
HAS PART: buffer#6, fender#4
HAS PART: bumper#2
HAS PART: bumper guard#1
HAS PART: car door#1
HAS PART: armrest#1
HAS PART: doorlock#1
HAS PART: hinge#1, flexible joint#1
HAS PART: pintle#1
wnb

• Derivationally related (different part-of-speech):

<table>
<thead>
<tr>
<th>2 of 8 senses of cat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense 3</td>
</tr>
<tr>
<td>cat#3</td>
</tr>
</tbody>
</table>
  RELATED TO->(adj) catty#1#1
  => bitchy#1, catty#1, cattish#1

| Sense 5              |
| cat-o'-nine-tails#1, cat#5 |
  RELATED TO->(verb) cat#1#1
  => cat#1
• Domain:

Sense 6
Caterpillar#2, cat#6
  USAGE-(noun) slang#2, cant#3, jargon#1, lingo#1, argot#1, patois#1, vernacular#1
  USAGE-(noun) trademark#2

Sense 1
cat#1, auto#1, automobile#1, machine#6, motorcar#1
  TOPIC TERM->(adj) hopped-up#1
  TOPIC TERM->(noun) rental#2, renting#1
  TOPIC TERM->(noun) alternator#1
  TOPIC TERM->(noun) backseat#2
  TOPIC TERM->(noun) road map#2
  TOPIC TERM->(noun) showroom#1, salesroom#1, saleroom#1
  TOPIC TERM->(noun) spark lever#1
  TOPIC TERM->(noun) tunnel#1
  TOPIC TERM->(noun) passenger#1, rider#4
  TOPIC TERM->(noun) grip#4, traction#1, adhesive friction#1
  TOPIC TERM->(noun) chattel#1, personal chattel#1, movable#1
  TOPIC TERM->(verb) prang#1
• Verbs

The verb destroy has 4 senses (first 3 from tagged texts)

1. (42) **destroy**#1, destruct#2 -- (do away with, cause the destruction or undoing of; "The fire destroyed the house")
2. (28) **destroy**#2, ruin#1 -- (destroy completely; damage irreparably; "You have ruined my car by pouring sugar in the tank!"; "The tears ruined her make-up")
3. (2) demolish#3, **destroy**#3 -- (defeat soundly; "The home team demolished the visitors")
4. **destroy**#4, put down#6 -- (put (an animal) to death; "The customs agents destroyed the dog that was found to be rabid"; "the sick cat had to be put down")