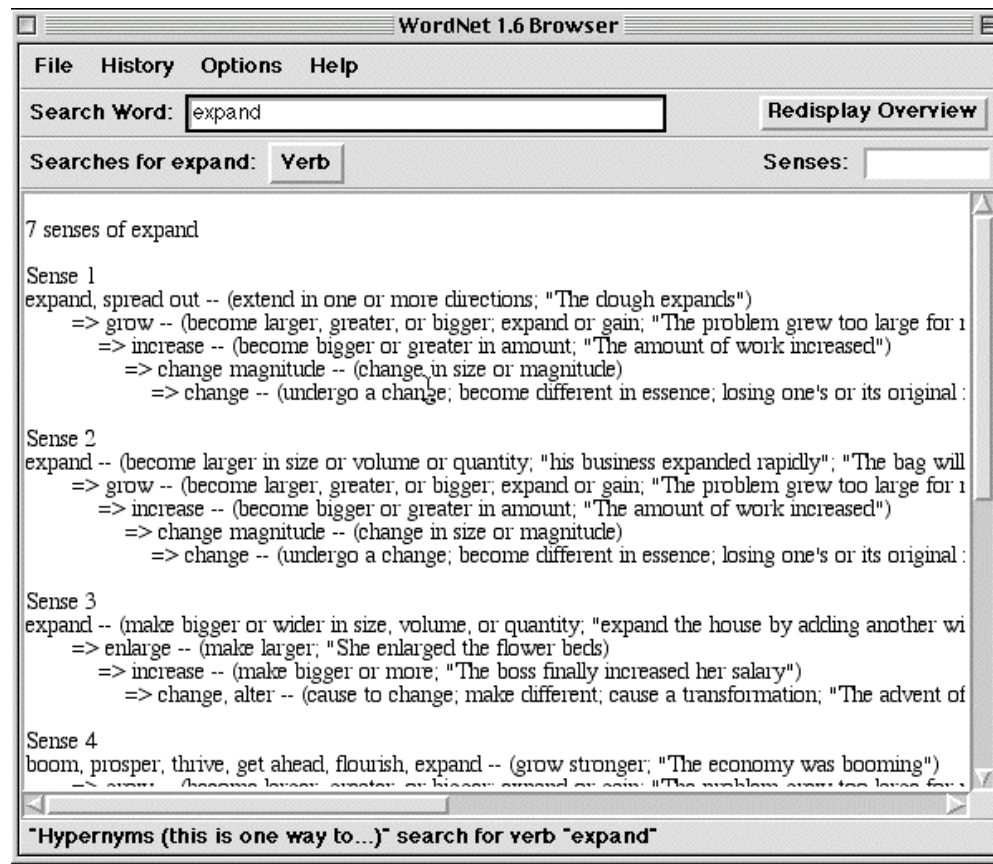


C SC 620
Advanced Topics in Natural
Language Processing

Lecture Notes 3

1/22/04

WordNet Browser Software



WordNet Browser Software

- For most platforms, see the Princeton website:
 - <http://www.cogsci.princeton.edu/~wn/>
- For MacOS X (*the one I've been using*), see:
 - <http://wordnet.sourceforge.net/>

wnconnect Software

- <http://linguistics.arizona.edu/~sandiway/wnconnect/>



What is it?

Given two terms, **wnconnect** is a program that finds and reports all possible connections between them in **WordNet**, a popular and freely-available synset (synonym set) network with semantic relations. WordNet is from Princeton University, see <http://www.cogsci.princeton.edu/~wn/>.

Example: *happy* and *sad*

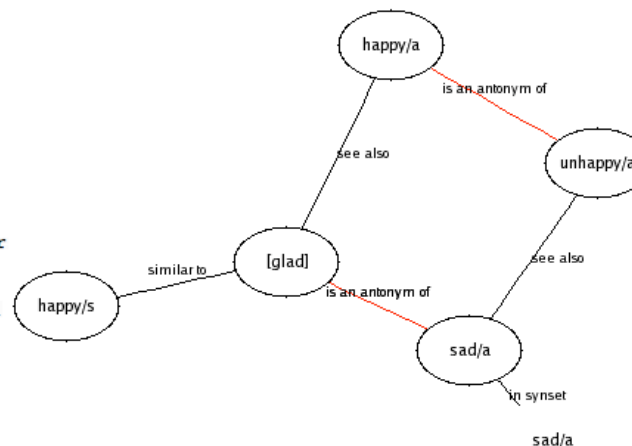
Here are the three shortest paths between the two terms, as reported by **wnconnect**. Synsets are enclosed in square brackets. The paths can be summarized graphically, as shown on the right:

```
happy/a is in [happy]
happy and unhappy/a are antonyms
For [unhappy], see also [sad]
sad/a is in the synset [sad]
```

```
happy/a is in [happy]
For [happy], see also [glad]
glad and sad/a are antonyms
```

```
happy/s is in [happy,pleased]
[happy,pleased] and [glad] are similar
glad and sad/a are antonyms
```

[There are actually 6 different paths between *happy* and *sad* in WordNet 1.7.1. The full graph generated by **wnconnect** is shown [here](#).]



wnconnect Software

- Does a breadth-first search to find shortest or all connections between two words
 - MacOS X version uses **neato** (from the freely available **Graphviz** package) to render graphs
- TTY-interface version now available for the Windows platform
 - (courtesy of TszYan Sandy Chow)
 - Uses SWI-Prolog
 - Free download from www.swi-prolog.org
 - Usage (Prolog syntax):
 - ?- connect(*happy,sad*).

wnconnect

File Edit View Favorites Tools Help

Address C:\Documents and Settings\Administrator.EMW\Desktop\wnconnect

wnconnect

connection Prolog Source

Modified: SWI-Prolog -- c:/Documents and Settings/Administrator.EMW/Desktop/wnconnect/connection.pl

Size: 17.6 KB

Attributes

```

File Edit Settings Run Debug Help
% ppl compiled 0.01 sec, 10,376 bytes
% sa compiled 0.05 sec, 266,212 bytes
% sim compiled 0.32 sec, 1,440,072 bytes
% vgp compiled 0.03 sec, 105,928 bytes
% s2 compiled 4.27 sec, 22,800,184 bytes
% s3 compiled 3.72 sec, 17,232,348 bytes
% s4 compiled 2.27 sec, 10,358,312 bytes
% file compiled 1.31 sec, 5,697,588 bytes
% fn compiled 0.85 sec, 4,246,372 bytes
% lexnames compiled 0.01 sec, 5,452 bytes
% ctr compiled 0.00 sec, 2,504 bytes
% c:/Documents and Settings/Administrator.EMW/Desktop/wnconnect/connection.pl compiled 18.19 sec, 95,159,340 bytes
Welcome to SWI-Prolog (Multi-threaded, Version 5.2.11)
Copyright (c) 1990-2003 University of Amsterdam.
SWI-Prolog comes with ABSOLUTELY NO WARRANTY. This is free software,
and you are welcome to redistribute it under certain conditions.
Please visit http://www.swi-prolog.org for details.

For help, use ?- help(Topic). or ?- apropos(Word).

1 ?- connect(happy,unhappy).

happy/a is in [euphoric, happy]
euphoric and dysphoric/a are antonyms
unhappy/a is in the synset [dysphoric, distressed, unhappy]
#nodes visited: 10

happy/a is in [happy]
happy and unhappy/a are antonyms
#nodes visited: 24

```

Type: Prolog Source Size: 17.6 KB

17.6 KB My Computer

Start wnconnect SWI-Prolog -- c:/Docu... Document1 - Microsoft W...

Prolog as a Database Query Language

- *Can use other tools, of course ...*
- Prolog as a logic-based database query language:
 - *Bust* as a verb belongs to 5 different synsets, i.e. has five different senses:

- `?- findall(S),s2(bust,v,_,S,_),L).L =
[200266721,201083468,201083844,201246161,201961102]`

- Three of these have the same file number 35 (verb.contact):

- `?-
findall((S,F),(s2(bust,v,_,S,_),file(S,F)),L).
L = [(200266721,30),(201083468,35),
(201083844,35),(201246161,35),(201961102,41)]`

Prolog as a Database Query Language

- The glosses corresponding to these three senses are:
 - `?- findall((S,G),(s2(bust,v,I,S,_),file(S,35),g(I,G)),L).`
 - `L =`
 - `[(201083468,'(go to pieces; "The lawn mower finally broke"; "The gears wore out"; "The old chair finally fell apart completely"'))],`
 - `(201083844,'(ruin completely; "He busted my radio!"))],`
 - `(201246161,'(separate or cause to separate abruptly; "The rope snapped"; "tear the paper"))]`
- The relation *common file number* will conflate these three senses.

Class Exercise

- What have (specific senses) of the following nouns in common?
 - *Umbrella*
 - *Saucepan*
 - *Baseball bat*
 - *Carpet beater*
- But do not share with:
 - *Giraffe*
 - *Pretzel*
 - *Homework*

Possible Software Project

- **Verb autoantonymy**

“Verbs that express both one meaning and an opposing meaning”

– **Example:** *dust*

– American Heritage Dictionary:

1.To remove dust ...

- **dust** the furniture

2.To apply dust ...

- **dust** the cookies with sugar

Autoantonyms

- Locate?
 - There is a website with a manually-compiled list of autoantonyms somewhere ...
- Possibly, it is this page on *antagonyms*:
 - <http://www-personal.umich.edu/~cellis/antagonym.html>

Antonyms

- Examples:
 - **Buckle:** to hold together (e.g. buckle your belt) vs. to fall apart (e.g., buckle under pressure) {AQ} **Bull:** A solemn edict or mandate vs. Nonsense or worthless information (3) {M}
 - **Chuff:** Elated vs. Unhappy (hinted at in 1) {M}
 - **Cite, Citation:** For doing good (such as military gallantry) vs. for doing bad (such as from a traffic policeman) (1)
 - **Cleave:** To adhere tightly vs. To cut apart (1) {A}
 - **Clip:** to attach vs. to cut off (1) {AH}{AS}
 - **Cool:** positive sense (cool web-sites) vs. negative sense(cool reception). {AA}