Computational Linguistics Lab: WordNet

Elena Cabrio, FBK-Irst
mail: cabrio@fbk.eu

Outline:

- Word sense identification basing on WordNet
  - Exercise of text annotation
- Compute semantic relatedness using lexical resources
  - WordNet::Similarity
- Reading group: discussion of the paper
  - Ponzetto, S. P. Strube, M. ***Knowledge Derived FromWikipedia For Computing Semantic Relatedness***

Word sense identification using WordNet

- Let's try to annotate a sample of text using WordNet synsets!

- Steps:
  - Identify the lemma of the content words
  - Look up into WordNet online **(http://wordnetweb.princeton.edu/perl/webwn)**
  - Disambiguate the sense of the word, given the context in which it appears
  - Annotate the word with the correct sense id

- Directory Lab4_exercise
  - plain_text.txt (sample of the Brown Corpus)
  - sample_to_annotate.txt

<s snum=29> The Protestants themselves are the first to admit the great falling off in effective membership in their churches. </s>

PLAIN TEXT

SAMPLE TO ANNOTATE

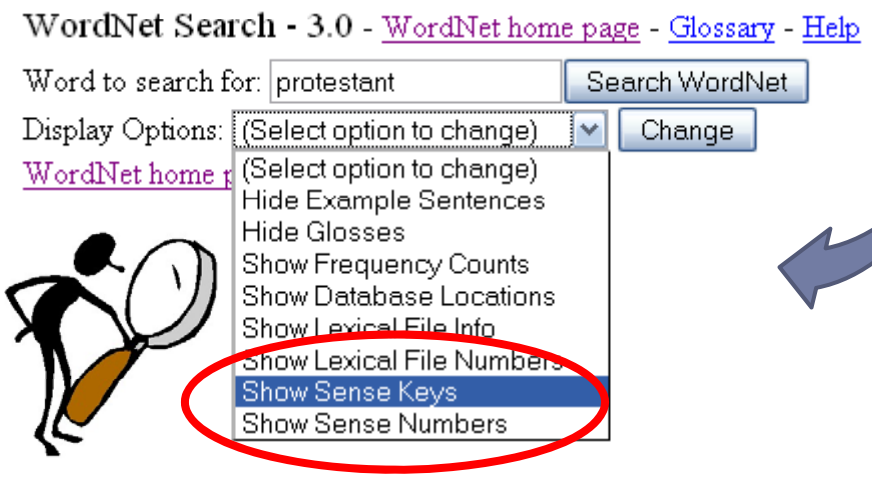
<s snum=29>
<wf cmd=ignore pos=DT>The</wf>
<wf cmd=done pos=NN lemma=? wnsn=? lexsn=? >Protestants</wf>
<wf cmd=ignore pos=PRP>themselves</wf>
<wf cmd=done pos=VB lemma= wnsn= lexsn= >are</wf>
…

<s snum=29> The Protestants themselves are the first to admit the great falling off in effective membership in their churches. </s>

PLAIN TEXT

SAMPLE TO ANNOTATE

<s snum=29>
<wf cmd=ignore pos=DT>The</wf>
<wf cmd=done pos=NN lemma=protestant wnsn=? lexsn=? >Protestants</wf>
<wf cmd=ignore pos=PRP>themselves</wf>
<wf cmd=done pos=VB lemma= wnsn= lexsn= >re</wf>
…



<s snum=29> The Protestants themselves are the first to admit the great falling off in effective membership in their churches. </s>

PLAIN TEXT

SAMPLE TO ANNOTATE

<s snum=29>
<wf cmd=ignore pos=DT>The</wf>
<wf cmd=done pos=NN lemma=protestant wnsn=? lexsn=? >Protestants</wf>
<wf cmd=ignore pos=PRP>themselves</wf>
<wf cmd=done pos=VB lemma= wnsn= lexsn= >are</wf>
…

WordNet Search - 3.0 - WordNet home page - Glossary - Help

Word to search for: protestant Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Noun

- S: (n) **Protestant#1 (protestant%1:18:00::)** (an adherent of Protestantism)
- S: (n) Protestant Church#1 (protestant_church%1:14:00::), **Protestant#2 (protestant%1:14:00::)** (the Protestant churches and denominations collectively)

Adjective

- S: (adj) **Protestant#1 (protestant%3:01:00::)** (of or relating to Protestants or Protestantism) *"Protestant churches"; "a Protestant denomination"*
- S: (adj) **protestant#2 (protestant%3:00:00:complaining:00)** (protesting)

<s snum=29> The Protestants themselves are the first to admit the great falling off in effective membership in their churches. </s>

PLAIN TEXT

SAMPLE TO ANNOTATE

<s snum=29>
<wf cmd=ignore pos=DT>The</wf>
<wf cmd=done pos=NN lemma=protestant wnsn=1 lexsn=1:18:00:: >Protestants</wf>
<wf cmd=ignore pos=PRP>themselves</wf>
<wf cmd=done pos=VB lemma= wnsn= lexsn= >are</wf>
…

WordNet Search - 3.0 - WordNet home page - Glossary - Help

Word to search for: protestant Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Noun

- S: (n) **Protestant#1 (protestant%1:18:00::)** (an adherent of Protestantism)
- S: (n) Protestant Church#1 (protestant_church%1:14:00::), **Protestant#2 (protestant%1:14:00::)** (the Protestant churches and denominations collectively)

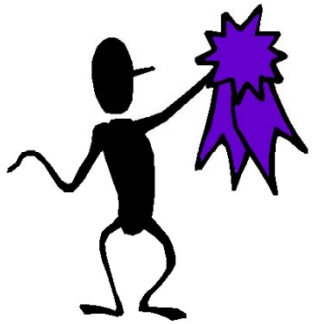
Adjective

- S: (adj) **Protestant#1 (protestant%3:01:00::)** (of or relating to Protestants or Protestantism) *"Protestant churches"; "a Protestant denomination"*
- S: (adj) **protestant#2 (protestant%3:00:00:complaining:00)** (protesting)

Evaluate your annotation

- Directory Lab4_exercise
  - goldstandard.txt (SemCor, sample of Brown Corpus annotated with WN synsets)
  - scorer.pl: to evaluate the accuracy of your work
    - From the command line:

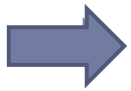
      perl scorer.pl sample_to_annotate.txt goldstandard.txt



Challenge for NLP systems: SENSEVAL
http://www.senseval.org/senseval3

Semantic similarity measures

- The need to determine the semantic distance between two lexical expressed concepts is a problem that pervades the Computational Linguistics
- The problem has a long history in philosophy, psychology, artificial intelligence, …



The perspective of semantic relatedness of two lexemes in a lexical resource

Word sense identification using WordNet

- In the CL literature, many terms are used by different authors: semantic relatedness, similarity, semantic distance, …
  - Resnik (1995): "car and gasoline seem more closely related than car and bicycles, even if the second pairs are more similar"
  - Morris and Hirst (2004): relatedness covers various kinds of relations: meronymy, antonymy, functional association, and other non-classical relations. E.g. antonyms are dissimilar and hence distant, but strongly related

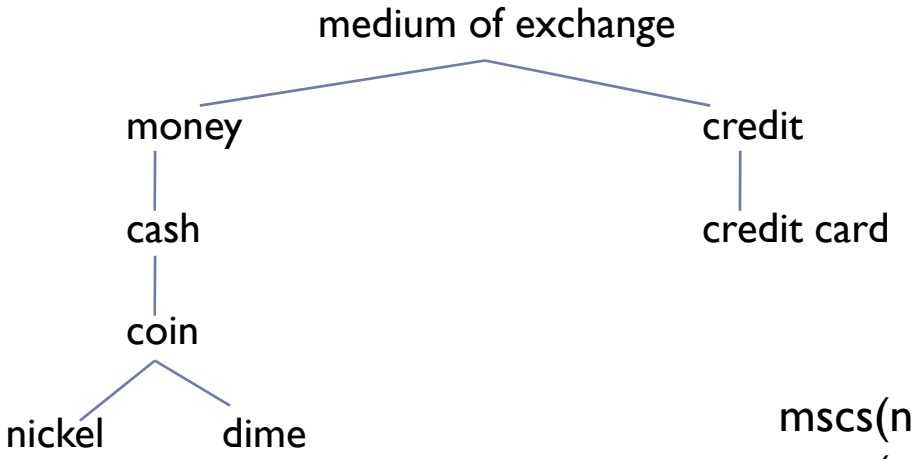
Semantic distance in a lexical resource

- **concept:** a particular sense of a given word
- we are not talking about similarity of distributional behavior of words (e.g. Dagan, 2000)
- when we refer to hierarchies and networks of concepts, we use the term **link** (or **edge**) to refer the relationships between nodes

- Approaches using WN are focused on noun hierarchy:
  - The backbone of noun hierarchy is the IS-A relation
  - At the top of the hierarchy are 11 abstract concepts (unique beginners)
  - The maximum depth is 16 nodes
  - Beyond IS-A, also other relations: Part-of, Substance-of, …

WordNet approaches: terminology

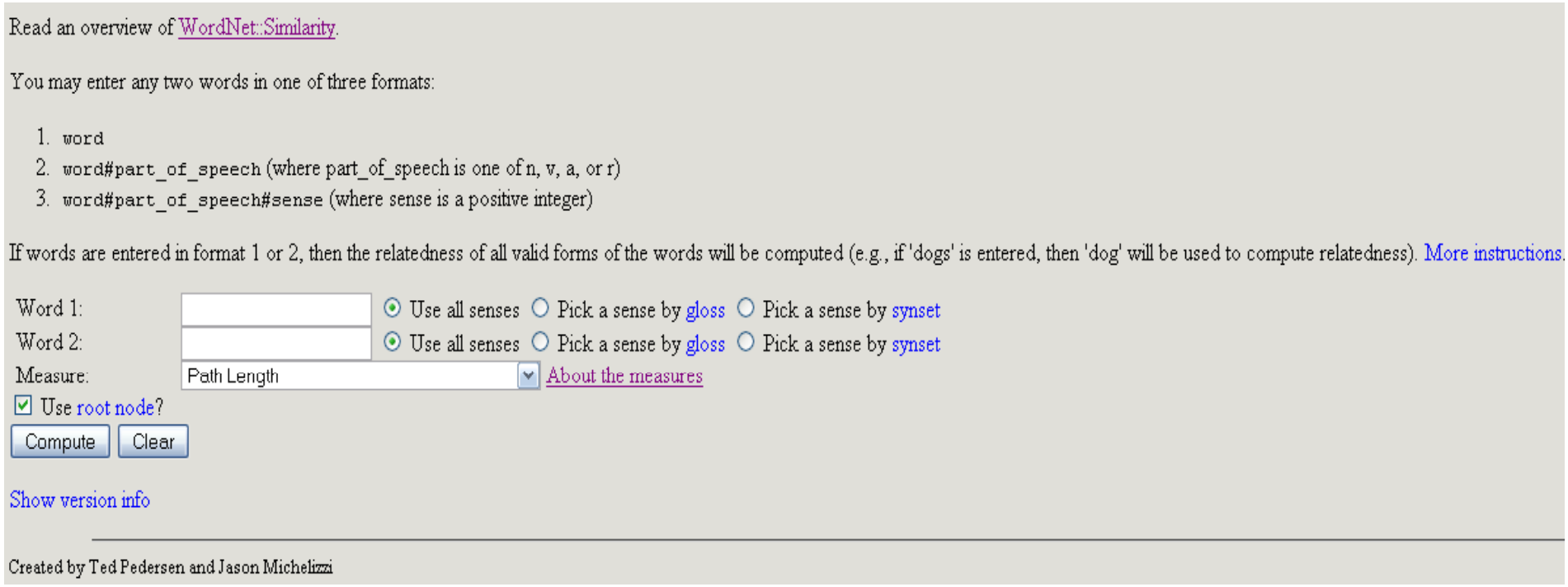
- The length of the ***shortest path*** form synset *ci* to *cj* (measured in edges) is denoted by ***len(ci,cj)***
- The ***depth of a node*** is the length from the node to the global root ***depth(ci) = len(root,ci)***
- ***mscs(ci,cj)*** is the most specific common subsumer of ci and cj



mscs(nickel,dime) = coin
mscs(nickel,credit_card) = medium_of_exchange

WordNet:: similarity

- http://marimba.d.umn.edu/cgi-bin/similarity/similarity.cgi



Reading group:

- Discussion of the paper:
  - Ponzetto, S. P. Strube, M. ***Knowledge Derived FromWikipedia For Computing Semantic Relatedness***

  http://www.jair.org/media/2308/live-2308-3485-jair.pdf

BIBLIOGRAPHY:

- Dagan, Ido. 2000. *Contextual Word Similarity*, in Rob Dale, Hermann Moisl and Harold Somers (Eds.), Handbook of Natural Language Processing, Marcel Dekker Inc, 2000, Chapter 19, pp. 459--476.
- Morris J. and Hirst G. (2004). *Non-classical lexical semantic relations*. In *Workshop on Computational Lexical Semantics, Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 46–51, May.
- Resnik, P. (1995) *Using Information Content to Evaluate Semantic Similarity in a Taxonomy*. International Joint Conference for Artificial Intelligence (IJCAI-95). 448-453.

- CREDITS: Thanks to Carlo Strapparava (FBK-Irst)