WordNet

How are words related?

- Word classes
  - Part of speech
- Selectional rules
  - ‘edible’ nouns
- Polarity - positive vs. negative words

Some relations

- Synonyms - ‘joy’, ‘elation’, ‘happiness’
- Antonyms - ‘love’, ‘hate’
- Homographs - ‘bank’, ‘bank’
- Hyponymy - ‘red’ is a hyponym of ‘color’
- Meronymy - ‘finger’ is a meronym of ‘hand’

Anything else?

- What sort of relationships would you like to know about between words?

WordNet

- Represents the relationship between Nouns, Verbs, Adjectives, Adverbs
- Many words have multiple senses.

Tree Senses

Noun

- S: (n) **tree** (a tall perennial woody plant having a main trunk and branches forming a distinct elevated crown; includes both gymnosperms and angiosperms)
- S: (n) **tree**, tree diagram (a figure that branches from a single root) "*genealogical tree*"
- S: (n) **Tree**, Sir Herbert Beerbohm Tree (English actor and theatrical producer noted for his lavish productions of Shakespeare (1853-1917))

Verb

- S: (v) corner, **tree** (force a person or an animal into a position from which he cannot escape)
- S: (v) **tree** (plant with trees) "*this lot should be treed so that the house will be shaded in summer*"
- S: (v) **tree** (chase an animal up a tree) "*the hunters treed the bear with dogs and killed it*"; "*her dog likes to tree squirrels*"
- S: (v) **tree**, shoetree (stretch (a shoe) on a shoetree)

Synsets

- Synsets, synonym sets, are a major component of WordNet.
- Antonyms are also included for some terms

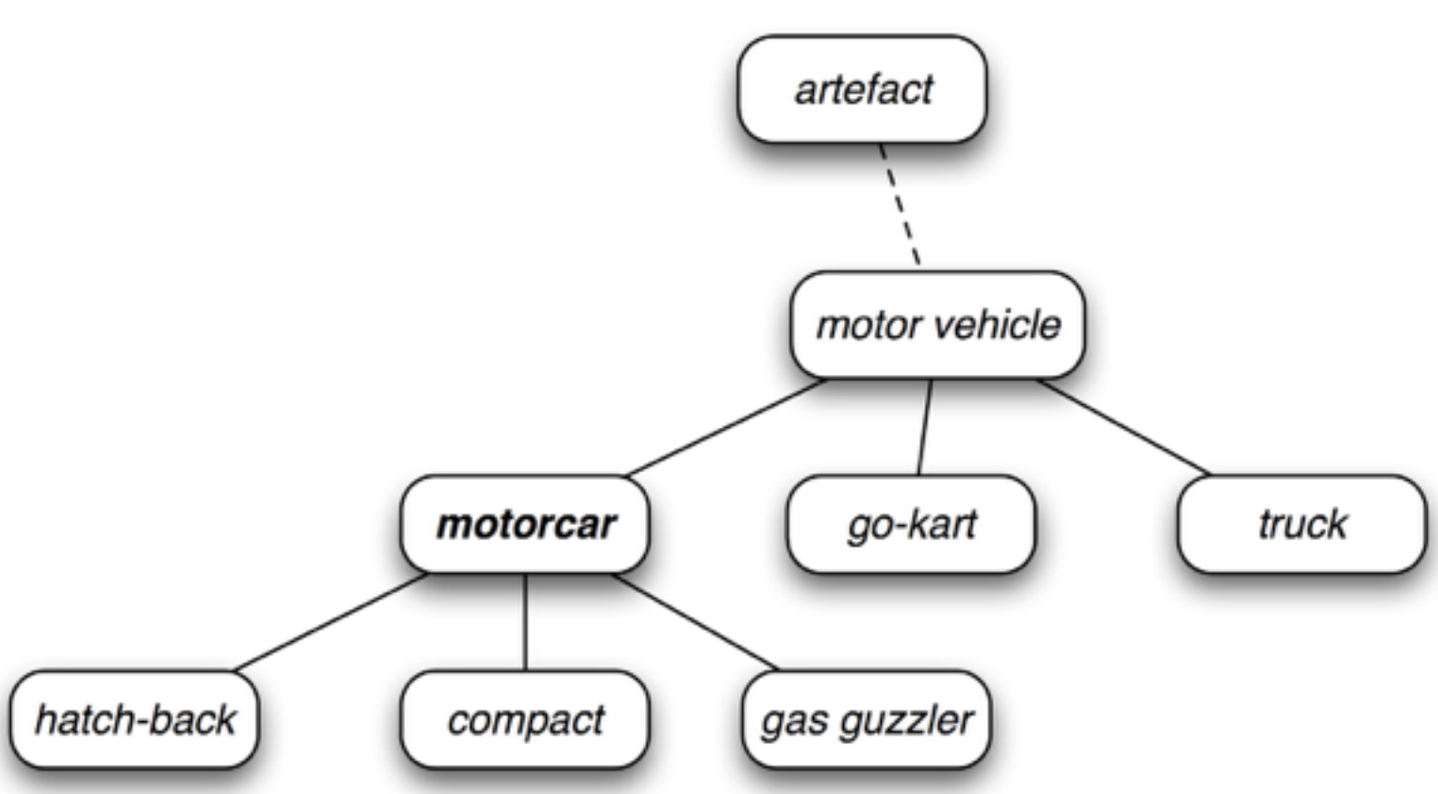
- S: (adj) **happy** (enjoying or showing or marked by joy or pleasure) *"a happy smile"; "spent many happy days on the beach"; "a happy marriage"*
  - *see also*
    - S: (adj) cheerful (being full of or promoting cheer; having or showing good spirits) *"her cheerful nature"; "a cheerful greeting"; "a cheerful room"; "as cheerful as anyone confined to a hospital bed could be"*
    - S: (adj) contented, content (satisfied or showing satisfaction with things as they are) *"a contented smile"*
    - S: (adj) glad (showing or causing joy and pleasure; especially made happy) *"glad you are here"; "glad that they succeeded"; "gave a glad shout"; "a glad smile"; "heard the glad news"; "a glad occasion"*
    - S: (adj) elated (exultantly proud and joyful; in high spirits) *"the elated winner"; "felt elated and excited"*
    - S: (adj) euphoric (exaggerated feeling of well-being or elation)
    - S: (adj) felicitous (exhibiting an agreeably appropriate manner or style) *"a felicitous speaker"*
    - S: (adj) joyful (full of or producing joy) *"make a joyful noise"; "a joyful occasion"*
    - S: (adj) joyous (full of or characterized by joy) *"felt a joyous abandon"; "joyous laughter"*

Other relationships in WordNet

- Hypernyms/Hyponyms
- Derivationally Derived Terms
- Examples of each sense
- Traditional Dictionary Style Definitions

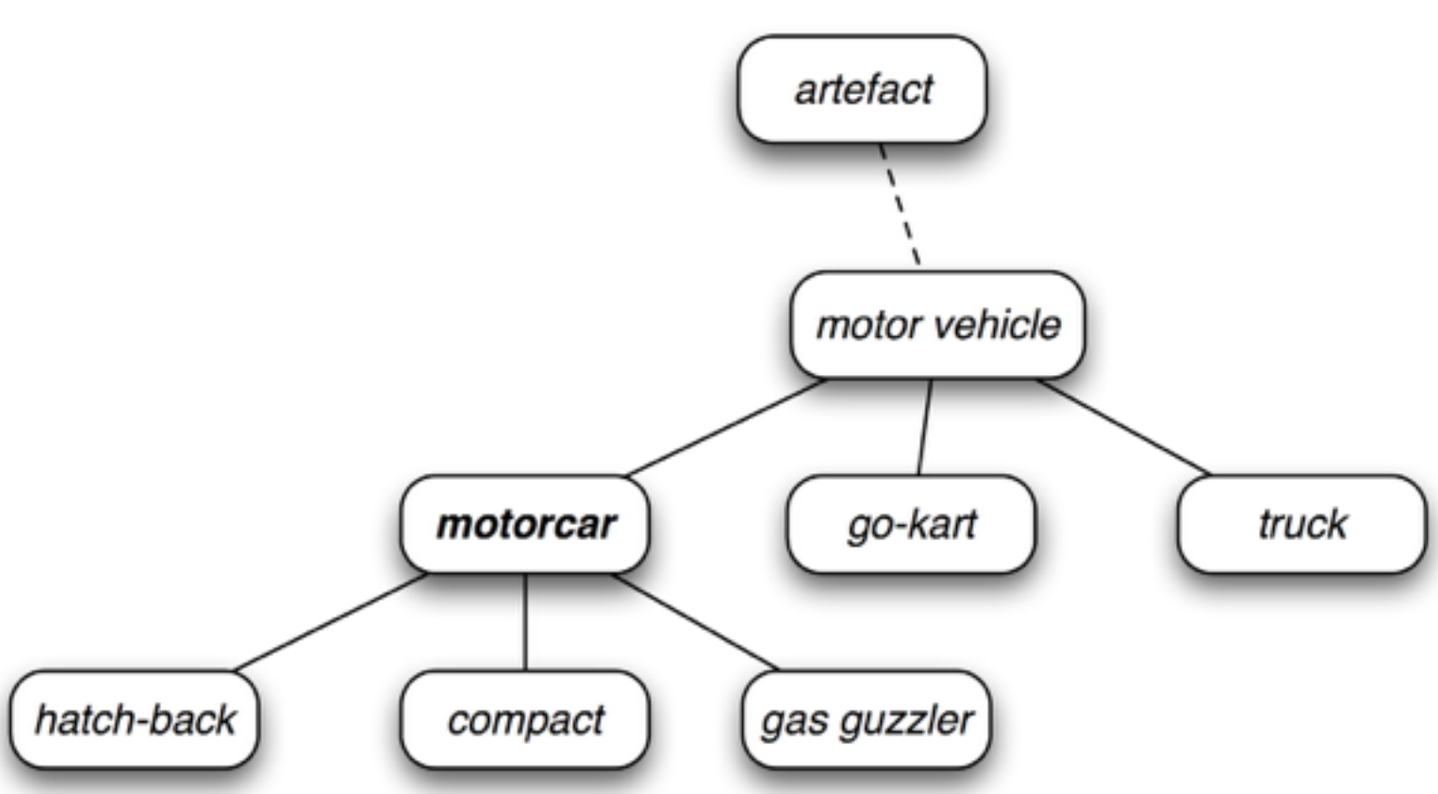
WordNet Similarity

- We can treat WordNet as a graph.
- Each sense is a node in the graph.
- Every relationship is an edge.



WordNet Similarity

- Similarity can be defined as the **distance** in the graph between two senses.
  - Distance is the number of edges between two nodes



WordNet Similarity

- Path length
  - sim(a,b) = -log pathlen(a,b)
- Resnik Similarity
  - sim(a,b) = -log p(LCS(a,b))
    - LCS = Lowest Common Subsumer
    - p(a) = the percentage of tokens (in a corpus) that are instances of a "concept"

Word Sense Disambiguation

- “I put my money in the bank.”
- “I slept on the bank of the river last night.”
- **bank** has multiple senses, and you’d like to know which to select.
- Choose the sense that is ‘closest’ to other words in the sentence.

Stemming and Lemmatizing

- It's valuable to be able to compare inflected forms of words.
  - walk = walks = walking
  - good = better = best
- Why not use regular expressions for this?
  - walking = walk
  - sing != s
  - strangest = strange
  - contest != cont

Normalization in NLTK

- Stemmers:
  - nltk.PorterStemmer()
  - nltk.LancasterStemmer()
- Lemmatizer - Each lemma is in the lexicon:
  - nltk.WordNetLemmatizer()

WordNet Demo

- Exploring WordNet
  - synsets, lemmas, definitions
  - common hypernyms
- Text Normalization
  - Stemming
  - Lemmatizing