

# NLP

# Text Similarity

*Semantic Similarity:  
Synonymy and other Semantic  
Relations*

# Synonyms and Paraphrases

- Example: post-close market announcements

The S&P 500	<u>climbed</u>	6.93, or 0.56 percent, to 1,243.72,	<u>its best close</u>	since June 12, 2001.
The Nasdaq	<u>gained</u>	12.22, or 0.56 percent, to 2,198.44	<u>for its best showing</u>	since June 8, 2001.
The DJIA	<u>rose</u>	68.46, or 0.64 percent, to 10,705.55,	<u>its highest level</u>	since March 15.

# Synonyms

- Different words (and also word compounds) can have similar meanings.
  - For example, the adjectives *tepid* and *lukewarm* have very similar meanings and can be substituted for one another (*tepid water* vs. *lukewarm water*).
- True synonyms are actually relatively rare.
  - For example, even though *big* and *large* are often thought of as synonyms, consider the difference between *Big Leagues* and *Large Leagues*. 😊
- The verbs *sweat* and *perspire* are also near synonyms.
  - However, they differ in their frequency of use and the type of text in which they are likely to appear.

## Polysemy

- Polysemy is the property of words to have multiple senses.
- For example, the noun *book* can refer to the following:
  - A literary work (e.g., “Anna Karenina”)
  - A stack of pages (e.g., a notebook)
  - A record of business transactions (think “bookkeeper”)
  - A record of bets (think “bookmaker”)
  - A list of buy and sell orders in a financial market

# Polysemy

- The same word can also have multiple parts of speech, each with its own set of senses. For example, the word *book*, as a verb can mean “make a reservation for” or “occupy”.
- The different senses of the same word don’t have to be equally frequent.
- Some of the senses may overlap (e.g., the first two senses of *book* on the previous slide). That’s partially why different dictionaries list different sets of word senses for the same word.
  - “My favorite books are Anna Karenina and my father’s checkbook” 😊
- Some words can be highly polysemous (e.g., the verb “get” has at least 35 different meanings, according to Wordnet).

## Other Semantic Relations

- Antonymy (near opposites)
  - *raise–lower*
- Hypernymy
  - a *deer* is a hypernym for *elk*
- Hyponymy (the inverse of hypernymy)
- Membership Meronymy:
  - a *flock* includes *sheep* (or *birds*)
- Part Meronymy:
  - a *table* has *legs*

## Synsets

- Semantic relations hold between word senses, not between words.
- Examples:
  - the antonym of *hot* can be either *mild* or *cold* (or *unattractive*) depending on the specific sense of *hot*.
  - the immediate hypernym of *bar* can be one of the following, among others: *room*, *musical notation*, *obstruction*, *profession*, depending on the sense of *bar*.
- The term *synset* is used to group together all synonyms of the same word. If a word is polysemous, it may be associated with multiple synsets.



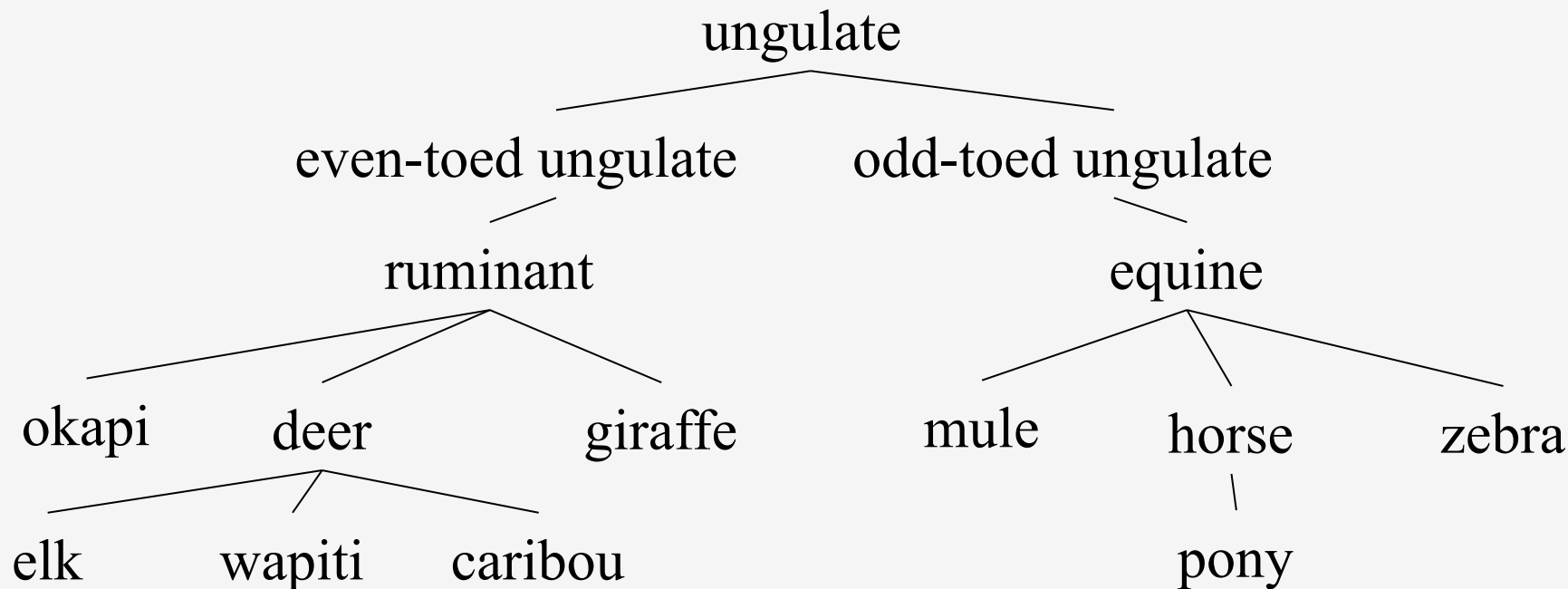
# Text Similarity

*Wordnet*

# Wordnet

- Wordnet is a project run by George Miller (1920–2012) and Christiane Fellbaum at Princeton University.
- It includes a database of words (mainly nouns and verbs but also adjectives and adverbs) and semantic relations between them.
- The main relation is hypernymy, so the overall structure of the database is more tree-like (see next slide).
- References:
  - George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39–41.
  - Christiane Fellbaum (1998, ed.) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

# Tree-like Structure of Wordnet



# Wordnet Example (1/6)

The noun bar has 11 senses

1. barroom, bar, saloon, ginmill, taproom -- (a room where alcoholic drinks are served over a counter)
2. bar -- (a counter where you can purchase food or drink)
3. bar -- (a rigid piece of metal)
4. measure, bar -- (notation for a repeating pattern of musical beats; written followed by a vertical bar)
5. bar -- (usually metal placed in windows to prevent escape)
6. prevention, bar -- (the act of preventing)
7. bar -- (a unit of pressure equal to a million dynes per square centimeter)
8. bar -- (a submerged (or partly submerged) ridge in a river or along a shore)
9. legal profession, bar, legal community -- (the body of individuals qualified to practice law)
10. cake, bar -- (a block of soap or wax)
11. bar -- ((law) a railing that encloses the part of the courtroom where the the judges and lawyers sit and the case is tried)

The verb bar has 4 senses

1. bar, debar, exclude -- (prevent from entering; keep out; "He was barred from membership in the club")
2. barricade, block, blockade, block off, block up, bar -- (render unsuitable for passage; "block the way"; "barricade the streets")
3. banish, relegate, bar -- (expel, as if by official decree; "he was banished from his own country")
4. bar -- (secure with, or as if with, bars; "He barred the door")

# Wordnet Example (2/6)

## Sense 1

barroom, bar, saloon, ginmill, taproom

=> room

=> area

=> structure, construction

=> artifact, artefact

=> object, physical object

=> entity, something

## Sense 2

bar

=> counter

=> table

=> furniture, piece of furniture, article of furniture

=> furnishings

=> instrumentality, instrumentation

=> artifact, artefact

=> object, physical object

=> entity, something

# Wordnet Example (3/6)

Sense 3

bar

- => implement
- => instrumentality, instrumentation
- => artifact, artefact
- => object, physical object
- => entity, something

Sense 4

measure, bar

- => musical notation
- => notation, notational system
- => writing, symbolic representation
- => written communication, written language
- => communication
- => social relation
- => relation
- => abstraction

# Wordnet Example (4/6)

Sense 5

bar

=> obstruction, impediment, impedimenta

=> structure, construction

=> artifact, artefact

=> object, physical object

=> entity, something

Sense 6

prevention, bar

=> hindrance, interference, interfering

=> act, human action, human activity

Sense 7

bar

=> pressure unit

=> unit of measurement, unit

=> definite quantity

=> measure, quantity, amount, quantum

=> abstraction

# Wordnet Example (5/6)

Sense 8

bar

=> ridge

=> natural elevation, elevation

=> geological formation, geology, formation

=> natural object

=> object, physical object

=> entity, something

=> barrier

=> mechanism

=> natural object

=> object, physical object

=> entity, something



# Wordnet Example (6/6)

Sense 9

legal profession, bar, legal community

=> profession, community

=> occupation, vocation, occupational group

=> body

=> gathering, assemblage

=> social group

=> group, grouping

Sense 10

cake, bar

=> block

=> artifact, artefact

=> object, physical object

=> entity, something

# Familiarity and Polysemy

board used as a noun is familiar (polysemy count = 9)

bird used as a noun is common (polysemy count = 5)

cat used as a noun is common (polysemy count = 7)

house used as a noun is familiar (polysemy count = 11)

information used as a noun is common (polysemy count = 5)

retrieval used as a noun is uncommon (polysemy count = 3)

serendipity used as a noun is very rare (polysemy count = 1)

# Text Similarity

*Other Lexical Networks*

## External Links

- EuroWordNet
  - <http://www.illc.uva.nl/EuroWordNet/>
- Open Thesaurus
  - <http://www.openthesaurus.de/>
- Freebase
  - <http://www.freebase.com>
- DBPedia
  - <http://www.dbpedia.org>
- BabelNet
  - <http://babelnet.org>
- Various thesauri

BabelNet 2.5 - A very large mu... +

babelnet.org/search.jsp?word=song&lang=EN

about • statistics • preferences

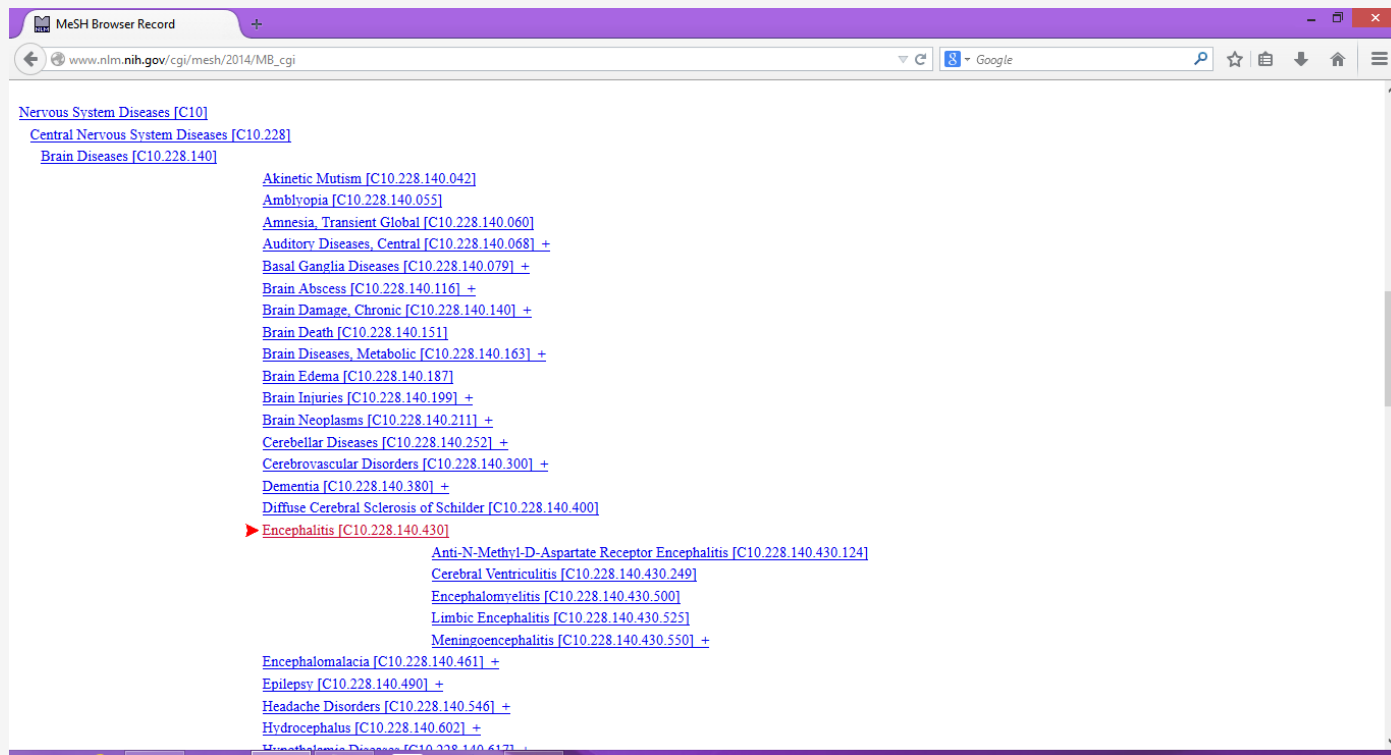
**Meaning:** song<sup>1</sup> • ID: bn:00072794n • Type: Concept [details] [explore]

**Senses:**

- [W] **song**<sup>1</sup>, vocal<sup>2</sup>  
 أغنية, 條, chanson, τραγούδι, canzone, canzonetta, シャンソン, シヤンゾン, ソング, 唱歌, 歌, 歌曲, 歌謡, 歌謡曲, 演歌, 謡物, canción
- [W] **Song**, 歌曲, Chanson, Lied, Τραγούδι, פֶּזֶם, गीत, Canzone (musica), 歌, Песня, Canción
- [W] Songs, Music track, 歌, Track (song), Canciones, الأغاني, الأغنية, أغني, ἀγαλλία, Ἀγάλλια, τραγουδιόματα, τραγουδιότρια, Ασµα, Brano, Brani musicali, 歌声, 声楽曲, 唄, Песен, Песньна, Песни, Cancion (musica), Canción (música), Cancion, Canción (musica), Cancion popular, Temas musicales, Canción popular, Canciones, Cancion (música)
- [U] **song**, الأغنية, 歌曲, chanson, Lied, τραγούδι, שיר, गीत, canzone, 歌, 歌曲, песня, canción
- [w] **song**  
**song**, أغنية, 歌曲, chanson, Lied, Τραγούδι, פֶּזֶם, गीत, canzone, 歌, песня, canción
- chanson, chansons, τραγούδι, τραγουδια, שירים, गीत, गाने, canzone, canzoni, песни, canción, canciones
- chanson, τραγούδι, गीत, canzone, песни, canción

**Glosses:** A short musical composition with words; "a successful musical must have at least three good songs"

## MeSH

Medical  
Subject  
Headings

- <http://www.nlm.nih.gov/mesh/MBrowser.html>

# NLP