

Word Meaning and Similarity

WordNet and other Online Thesauri



Applications of Thesauri and Ontologies

- Information Extraction
- Information Retrieval
- Question Answering
- Bioinformatics and Medical Informatics
- Machine Translation



WordNet 3.0

- A hierarchically organized lexical database
- On-line thesaurus + aspects of a dictionary
 - Some other languages available or under development
 - (Arabic, Finnish, German, Portuguese...)

Category	Unique Strings
Noun	117,798
Verb	11,529
Adjective	22,479
Adverb	4,481



Senses of “bass” in Wordnet

Noun

- **S: (n) bass** (the lowest part of the musical range)
- **S: (n) bass**, **bass part** (the lowest part in polyphonic music)
- **S: (n) bass**, **basso** (an adult male singer with the lowest voice)
- **S: (n) sea bass**, **bass** (the lean flesh of a saltwater fish of the family Serranidae)
- **S: (n) freshwater bass**, **bass** (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
- **S: (n) bass**, **bass voice**, **basso** (the lowest adult male singing voice)
- **S: (n) bass** (the member with the lowest range of a family of musical instruments)
- **S: (n) bass** (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Adjective

- **S: (adj) bass**, **deep** (having or denoting a low vocal or instrumental range) *"a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"*



How is “sense” defined in WordNet?

- The **synset (synonym set)**, the set of near-synonyms, instantiates a sense or concept, with a **gloss**
- Example: **chump** as a noun with the **gloss**:
“a person who is gullible and easy to take advantage of”
- This sense of “chump” is shared by 9 words:
chump¹, fool², gull¹, mark⁹, patsy¹, fall guy¹,
sucker¹, soft touch¹, mug²
- Each of **these** senses have this same gloss



WordNet Hypernym Hierarchy for “bass”

- S: (n) bass, basso (an adult male singer with the lowest voice)
 - direct hypernym / inherited hypernym / sister term
 - S: (n) singer, vocalist, vocalizer, vocaliser (a person who sings)
 - S: (n) musician, instrumentalist, player (someone who plays a musical instrument (as a profession))
 - S: (n) performer, performing artist (an entertainer who performs a dramatic or musical work for an audience)
 - S: (n) entertainer (a person who tries to please or amuse)
 - S: (n) person, individual, someone, somebody, mortal, soul (a human being) *"there was too much for one person to do"*
 - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
 - S: (n) living thing, animate thing (a living (or once living) entity)
 - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
 - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
 - S: (n) physical entity (an entity that has physical existence)
 - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))



WordNet Noun Relations

Relation	Also called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	<i>breakfast</i> ¹ → <i>meal</i> ¹
Hyponym	Subordinate	From concepts to subtypes	<i>meal</i> ¹ → <i>lunch</i> ¹
Member Meronym	Has-Member	From groups to their members	<i>faculty</i> ² → <i>professor</i> ¹
Has-Instance		From concepts to instances of the concept	<i>composer</i> ¹ → <i>Bach</i> ¹
Instance		From instances to their concepts	<i>Austen</i> ¹ → <i>author</i> ¹
Member Holonym	Member-Of	From members to their groups	<i>copilot</i> ¹ → <i>crew</i> ¹
Part Meronym	Has-Part	From wholes to parts	<i>table</i> ² → <i>leg</i> ³
Part Holonym	Part-Of	From parts to wholes	<i>course</i> ⁷ → <i>meal</i> ¹
Antonym		Opposites	<i>leader</i> ¹ → <i>follower</i> ¹



WordNet 3.0

- Where it is:
 - <http://wordnetweb.princeton.edu/perl/webwn>
- Libraries
 - Python: WordNet from NLTK
 - <http://www.nltk.org/Home>
 - Java:
 - JWNL, extJWNL on sourceforge



MeSH: Medical Subject Headings thesaurus from the National Library of Medicine

- **MeSH (Medical Subject Headings)**
 - 177,000 entry terms that correspond to 26,142 biomedical “headings”

- **Hemoglobins**

Entry Terms: Eryhem, Ferrous Hemoglobin,
Hemoglobin

Definition: The oxygen-carrying proteins of ERYTHROCYTES. They are found in all vertebrates and some invertebrates. The number of globin subunits in the hemoglobin quaternary structure differs between species. Structures range from monomeric to a variety

Synset



The MeSH Hierarchy

1. + **Anatomy [A]**
2. + **Organisms [B]**
3. + **Diseases [C]**
4. - **Chemicals and Drugs [D]**
 - [Inorganic Chemicals \[D01\]](#) +
 - [Organic Chemicals \[D02\]](#) +
 - [Heterocyclic Compounds \[D03\]](#) +
 - [Polycyclic Compounds \[D04\]](#) +
 - [Macromolecular Substances \[D05\]](#) +
 - [Hormones, Hormone Substitutes, and](#)
 - [Enzymes and Coenzymes \[D08\]](#) +
 - [Carbohydrates \[D09\]](#) +
 - [Lipids \[D10\]](#) +
 - [Amino Acids, Peptides, and Proteins](#)
 - [Nucleic Acids, Nucleotides, and Nucl](#)
 - [Complex Mixtures \[D20\]](#) +
 - [Biological Factors \[D23\]](#) +
 - [Biomedical and Dental Materials \[D25\]](#) +
 - [Pharmaceutical Preparations \[D26\]](#) +

Amino Acids, Peptides, and Proteins [D12]

Proteins [D12.776]

Blood Proteins [D12.776.124]

Acute-Phase Proteins [D12.776.124.050] +

Anion Exchange Protein 1, Erythrocyte [D12.776.124.078]

Ankyrins [D12.776.124.080]

beta 2-Glycoprotein I [D12.776.124.117]

Blood Coagulation Factors [D12.776.124.125] +

Cholesterol Ester Transfer Proteins [D12.776.124.197]

Fibrin [D12.776.124.270] +

Glycophorin [D12.776.124.300]

Hemocyanin [D12.776.124.337]

➤ [Hemoglobins \[D12.776.124.400\]](#)

Carboxyhemoglobin [D12.776.124.400.141]

Erythrocyte proteins [D12.776.124.400.220]



Uses of the MeSH Ontology

- Provide synonyms (“entry terms”)
 - E.g., glucose and dextrose
- Provide hypernyms (from the hierarchy)
 - E.g., glucose ISA monosaccharide
- Indexing in MEDLINE/PubMED database
 - NLM’s bibliographic database:
 - 20 million journal articles
 - Each article hand-assigned 10-20 MeSH terms

Word Meaning and Similarity

WordNet and other Online Thesauri