



Glossary

Binary polarity

Assigns a ± 1 value to a document to indicate a positive or negative sentiment. In this case, a value of 0 indicates neutrality, or a lack of sentiment.

Conditional random fields (CRF) model

A powerful machine learning technique which takes advantage of the neighboring structure of a sequence of words.

Homographs

Words that are spelled identically but have different senses. “Bat,” for example, refers both to a club used for hitting a ball and a nocturnal flying animal.

Homonyms

A superset containing both homographs and homophones.

Homophones

Words that sound the same, but are spelled differently and have different senses. For example, the words “write” and “right” are homophones.

Hypernym

A general term that also includes the meaning of more specific terms. For example, a vehicle is a hypernym of a car, which is a hyponym of a vehicle in this parent-child relationship.

Hyponym

A word with a specific meaning that is part of a more general term. For example, a vehicle is a hypernym of a car, which is a hyponym of a vehicle in this parent-child relationship.

Holonym

The whole in a part/whole relationship. For example, lettuce and tomatoes are meronyms of a sandwich, which is their holonym.



Entailment

A verb causally evolved from some other verb. For example, the word “eat” causes the entailment “chew,” but not the other way around. Entailments allow for constructing or assessing a sequence of logical actions.

Lemma

A canonical form of a word with a single meaning. Several lemmas make up a synset.

Lexical database

Also, lexical ontologies. Databases such as WordNet, which contain semantic relationships and meanings that curators consider important for a set of NLP tasks. Often, such ontologists represent words and phrases as nodes of a tree graph.

Lexical taxonomy

Also, taxonomic trees. A tree graph representing a lexical database.

Lexicon

A vocabulary.

Meronymy

The part in a part/whole relationship. For example, lettuce and tomatoes are meronyms of a sandwich, which is their holonym.

Named entity (NE)

A word or a phrase which represents real-world objects, such as a person, an organization, a country or location, an event, and many others. It is an instance of an entity and it is often denoted with a proper noun. For example, New York City is an instance of a city. Bill Gates is an instance of a person. Microsoft is an instance of an organization.

Named entity recognition (NER)

Also, named entity extraction (NEE). A technique for recognizing named entities in a text.

Semantic analysis

A broad description of techniques focused on extracting, representing, and storing meaning of texts.

Sense

In WordNet, sense refers to one aspect, or meaning, of a word.



Sentiment

A subjective opinion, feeling, or polarity towards an event, product, or service.

Sentiment analysis

Also, opinion mining. The process of systematically classifying documents according to their polarities.

Synset

A set of synonymous lemmas, or synonyms which are semantically equivalent in some contexts.

Valence

A continuous numerical value indicating the sentiment of a text. This value is often normalized to a finite range, such as $[-1,1]$.

Valence Aware Dictionary for sEntiment Reasoning (VADER)

A successful lexicon-based model that aims to rank sentiment for shorter messages, including tweets and social media posts. Its established lexicon includes emoticons, acronyms, and slang.

