

WORDNET

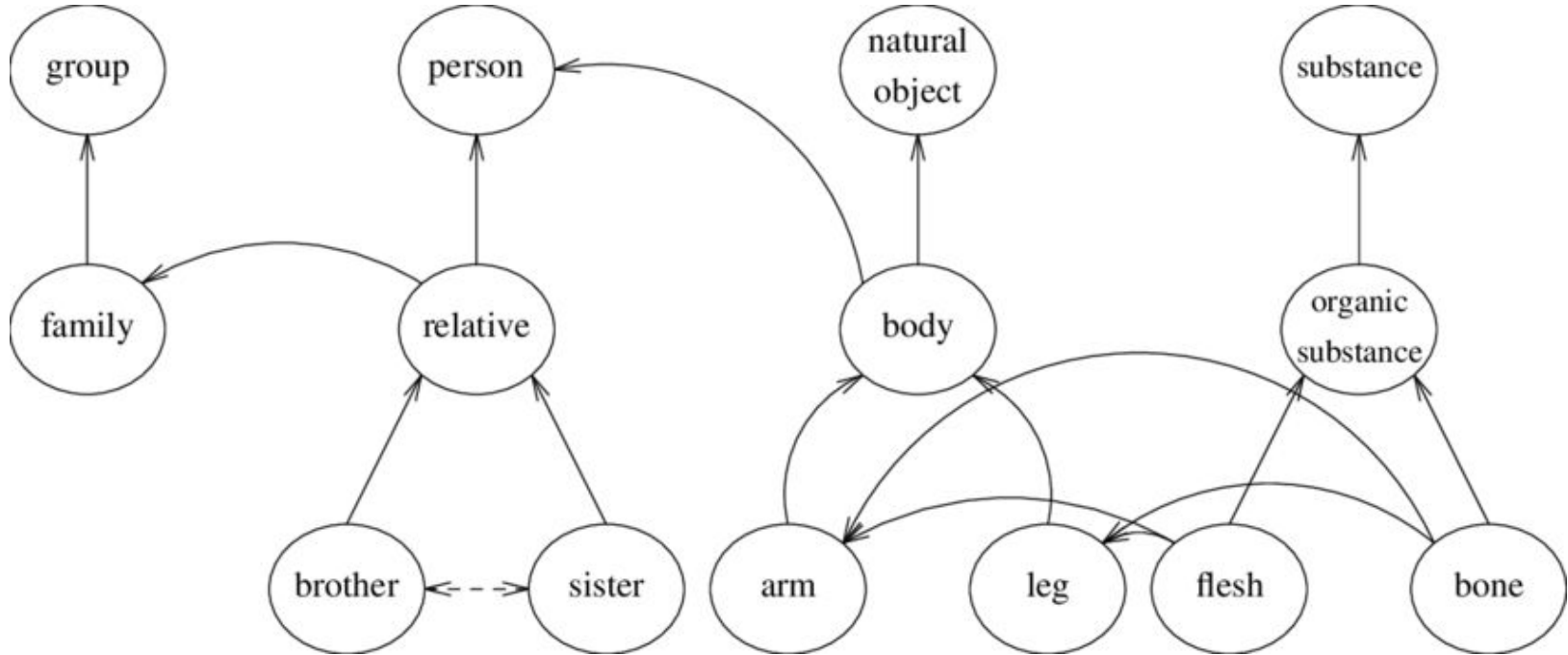
Bingyu Xiong
Profilierungsmodul Computerlinguistik I
Dr. Robert Zangenfeind

16.01.2024

ÜBERSICHT

- 1 Was ist WordNet
- 2 Synset
- 2 Substantive in WordNet
- 3 Adjektive in WordNet
- 4 Verben in WordNet
- 6 Anwendungen
- 7 EuroNet und GermaNet
- 8 WordNet im Vergleich zu BTM
- 9 Zusammenfassung

WAS IST WORDNET?



Quelle: Miller, George & Beckwith, R. & Fellbaum, Christiane & Gross, Derek & Miller, Katherine. (1991). Introduction to WordNet: An On-line Lexical Database*. 3. 10.1093/ijl/3.4.235.

KERNKONZEPT: SYNSET

- ❑ Die Hauptbeziehung zwischen Wörtern in WordNet sind Synonyme
- ❑ Einträge werden nach ihren Bedeutungen gruppiert, jede Gruppe mit derselben Bedeutung wird als Synset bezeichnet
- ❑ WordNet bietet eine kurze, zusammenfassende Definition für jedes Synset an und zeichnet die semantischen Beziehungen zwischen verschiedenen Synsets auf

Word to search for: Display Options:

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

Display options for sense: (gloss) "an example sentence"

Noun

- [S:](#) (n) **read** (something that is read) "*the article was a very good read*"

Verb

- [S:](#) (v) **read** (interpret something that is written or printed) "*read the advertisement*"; "*Have you read Salman Rushdie?*"
- [S:](#) (v) **read**, [say](#) (have or contain a certain wording or form) "*The passage reads as follows*"; "*What does the law say?*"
- [S:](#) (v) **read** (look at, interpret, and say out loud something that is written or printed) "*The King will read the proclamation at noon*"
- [S:](#) (v) **read**, [scan](#) (obtain data from magnetic tapes or other digital sources) "*This dictionary can be read by the computer*"
- [S:](#) (v) **read** (interpret the significance of, as of palms, tea leaves, intestines, the sky; also of human behavior) "*She read the sky and predicted rain*"; "*I can't read his strange behavior*"; "*The fortune teller read his fate in the crystal ball*"
- [S:](#) (v) [take](#), **read** (interpret something in a certain way; convey a particular meaning or impression) "*I read this address as a satire*"; "*How should I take this message?*"
- [S:](#) (v) [learn](#), [study](#), **read**, [take](#) (be a student of a certain subject) "*She is reading for the bar exam*"
- [S:](#) (v) **read**, [register](#), [show](#), [record](#) (indicate a certain reading; of gauges and instruments) "*The thermometer showed thirteen degrees below zero*"; "*The gauge read 'empty'*"
- [S:](#) (v) **read** (audition for a stage role by reading parts of a role) "*He is auditioning for 'Julius Caesar' at Stratford this year*"
- [S:](#) (v) **read** (to hear and understand) "*I read you loud and clear!*"
- [S:](#) (v) [understand](#), **read**, [interpret](#), [translate](#) (make sense of a language) "*She understands French*"; "*Can you read Greek?*"

EINE KURZE ENTWICKLUNGSGESCHICHTE (1985-)

Hintergrund:

- ❑ 1970er Jahre: Lexikalische Semantik basierend auf Bedeutung der einzelnen lexikalischen Elemente
- ❑ 1980er Jahre: Beziehungsbasierte lexikalische Semantik

1978: George Miller (Cognitive Science Laboratory, Princeton University),
Automated dictionary

1985: George Miller, *Wordnet: A Dictionary Browser*

Frühjahr 1987: Johnson-Laird übernahm die Nomen Klassifizierung

Sommer 1987: Fellbaum fuhr die Verben Klassifizierung durch

Kitty Miller war für die Adjektiv Klassifizierung verantwortlich

WORDNETS VERSION UND UMFANG

- ❑ Die neueste Windows-Version von WordNet ist 2.1(März 2005)
- ❑ Version 3.0 für Unix/Linux/Solaris usw. wurde im Dezember 2006 veröffentlicht
- ❑ Version 3.1 derzeit nur online verfügbar
- ❑ Die Datenbank enthält 155.327 Wörter, organisiert in 175.979 Synsets für insgesamt 207.016 Wort-Sinn-Paare

WAS GIBT'S IN WORDNET

❑ Inhalte von WordNet:

zusammengesetzte Wörter, Phrasen, Kollokationen, Idiome und Wörter, wobei Wort die grundlegendste Einheit ist

❑ Semantische Beziehungen zwischen Objekten:

Synonyme und Antonyme

Oberbegriffe und Unterbegriffe

Beziehung zwischen Ganzem und Teil

...

BEZIEHUNGSSYMBOLLE

Nomen		Verb		Adjektive		Adverbien	
Antonym (Gegenwort)	!	Antonym	!	Antonym	!	Antonym	!
Hyponym (Unterbegriff)	~	Hyponym	~	Synonyme	&	Abgeleitet von	\
Hyperonym (Oberbegriff)	@	Hyperonym	@	relationales Adjektiv	\		
Meronym ("ist Teil von")	#	Konsequenz	*	Siehe auch	^		
Holonym ("ist Ganzes von")	%	Ursache	>	Attribut	=		
Attribut	=	Siehe auch	^				

SUBSTANTIVE IN WORDNET

- ❑ Synset und lexikalische Hierarchie

{robin, redbreast} @-> {animal, animal_being} @->
{organism, life_form, living_thing}

- ❑ 25 Grundkategorien

{act, activity} {food} {possession} {animal, fauna}
{group, grouping} {process} {artifact} ...

- ❑ Es gibt selten mehr als 10-12 Ebenen im semantischen Baum

shetland pony @-> pony @-> horse @-> equid @-> odd-toed ungulate @->
placental mammal @-> mammal @-> vertebrate @-> chordate @-> animal
@-> organism @-> entity

GANZ-TEIL-BEZIEHUNG (MERONYMIE)

- Es gibt die folgenden drei Situationen:

A ist Bestandteil von B (wing -> bird)

A ist Mitglied von B (tree -> forest)

A ist Material von B (aluminum -> plane)

- Problem: Teilnamen zu weit oben in der Hierarchie platziert

{wheel} ist ein Teil von {vehicle}

{sled} ist eine Art von {vehicle}

{wheel} ist NICHT ein Teil von {sled}

-> Zwischenkonzept {wheeled_vehicle}

ADJEKTIVE IN WORDNET

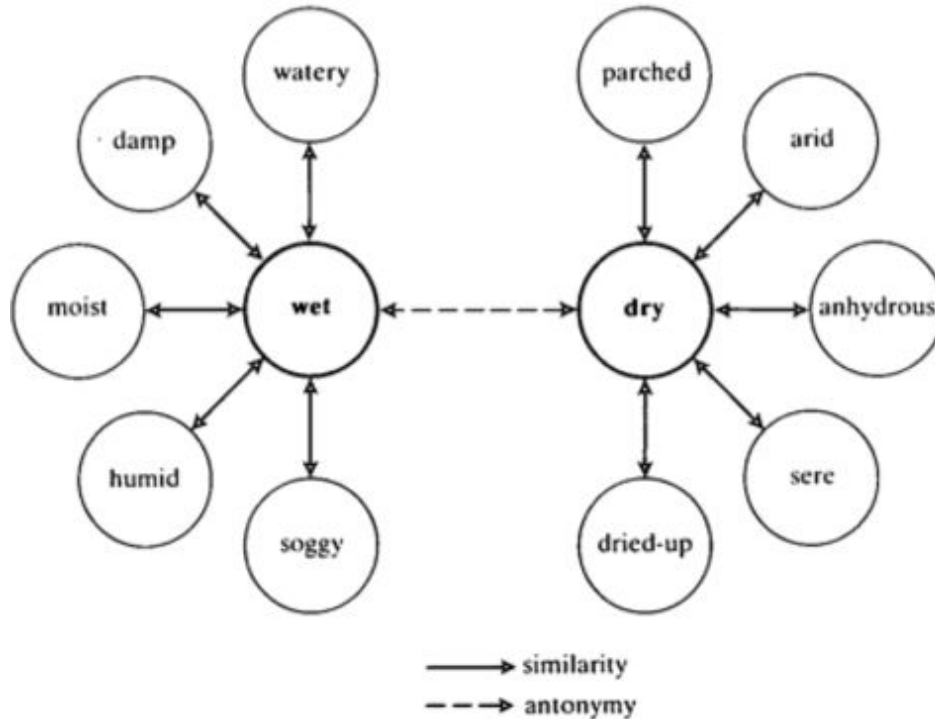
- Beschreibende Adjektive

z.B. groß, schön, interessant, möglich, verheiratet, ...

- Relationale Adjektive

z.B. fraternal, electrical, sidereal, ...

ANTONYME BESCHREIBENDER ADJEKTIVE



moist &-> wet !-> dry

MERKMALE RELATIONALER ADJEKTIVE

- können nur in attributiven Positionen erscheinen
- Bedeutung ist eng mit einem Substantiv verwandt
fraternal twins – fraternal: brother
dental hygiene – dental: tooth
- können nicht durch Gradadverbien modifiziert werden
the very baseball game(x)
- haben keine direkten Antonyme
civil lawyer vs. criminal lawyer

VERBEN IN WORDNET

❑ Klassifizierung englischer Verben

Lyons(1977): act, move, get, become, be, make

Pulman(1983): be, do = Stative Verben, Aktivität

Jackendoff(1983): event, state

❑ 15 grundlegende Kategorien von Wordnet-Verben :

motion	perception	contact
communication	competition	change
cognition	consumption	creation
emotion	stative	possession
body	social	weather

ANTONYM VON VERBEN IM WORDNET

❑ *give/take; buy/sell; lend/borrow; teach/learn*

-> Haben keine gemeinsame Hyperonyme, werden in derselben Szene verwendet

❑ *live/die; exclude/include; differ/equal; wake/sleep*

-> Verben, die einen Zustand ausdrücken

❑ *lengthen/shorten; strengthen/weaken; prettify/uglify*

-> Verben, die eine Veränderung ausdrücken

❑ *tie/untie; appear/disappear*

-> Widerspruch von Markierung

INDEX DER VERTRAUTHEIT

Table 4.4
Hypernyms of *bronco* and their index values

Word	Polysemy
bronco	1
@ → mustang	1
@ → pony	5
@ → horse	14
@ → equine	0
@ → odd-toed ungulate	0
@ → placental mammal	0
@ → mammal	1
@ → vertebrate	1
@ → chordate	1
@ → animal	4
@ → organism	2
@ → entity	3

Quelle: Christiane
Fellbaum(1998).WordNet: An
Electronic Lexical
Database.The MIT Press.

ANWENDUNGEN VON WORDNET

- ❑ Begriffsklärung
- ❑ Statistische Modelle basierend auf der Klassifizierung der Wortbedeutung
- ❑ Textsuche
- ❑ Korrekturlesen

...

VERWENDUNG VON WORDNET ZUM KORREKTURLESEN VON TEXTEN

Prinzip

- ❑ Kohärenter Text wird als lexikalische Ketten verkörpert
- ❑ Anhand der verschiedenen Beziehungen zwischen Wörtern in WordNet können die tatsächlichen lexikalischen Ketten berechnet werden, die zwischen Wörtern im Text gebildet werden
- ❑ Wenn es Wörter gibt, die nicht in die lexikalischen Ketten fallen, kann es sich um einen Fehler handeln

Ein Beispiel:

- ❑ much of that data, he notes, is available **toady**/today electronically.

Quelle: G.Hirst & D. St-Onge, 1998, Lexical Chains as Representation of Context for the Detection and Correction of Malapropism

EURONET & GERMANET

Euro-Wordnet

- ❑ 1996–1999
- ❑ Department of Computational Linguistics, University of Amsterdam
- ❑ Niederländisch, Italienisch, Spanisch, Deutsch, Französisch, Tschechisch und Estnisch

GermaNet

- ❑ Universität Tübingen
- ❑ 154.814 lexikalische Einheiten, ca. 120.032 Synonymenmengen (Synsets)

Quelle: Claudia Kunze, Semantische Relationstypen in GermaNet, in: Langer/Schnorbusch, Semantik im Lexikon (2005), S. 161–178.

IM VERGLEICH ZUM ERKLÄREND-KOMBINATORISCHEN WÖRTERBUCH IM BTM

- ❑ Beide Ansätze versuchen, die Bedeutung von Wörtern auf einer tieferen semantischen Ebene zu erfassen
- ❑ WordNet enthält keine Informationen zur Etymologie oder Aussprache von Wörtern und nur begrenzte Informationen zur Verwendung
- ❑ WordNet zielt darauf ab, die meisten alltäglichen Wörter abzudecken und enthält nicht viele domänenspezifische Terminologie

ZUSAMMENFASSUNG

- ❑ WordNet: eine lexikalische Datenbank semantischer Beziehungen zwischen Wörtern
- ❑ verknüpft Wörter in semantischen Beziehungen
- ❑ Synset
- ❑ Einschränkung: Einige der semantischen Beziehungen eignen sich eher für konkrete Konzepte als für abstrakte Konzepte

QUELLEN:

1. Miller, George & Beckwith, R. & Fellbaum, Christiane & Gross, Derek & Miller, Katherine. (1991). Introduction to WordNet: An On-line Lexical Database. 3. 10.1093/ijl/3.4.235.
2. Christiane Fellbaum(1998).WordNet: An Electronic Lexical Database.The MIT Press.
3. G.Hirst & D. St-Onge, 1998, Lexical Chains as Representation of Context for the Detection and Correction of Malapropism.
4. George A. Miller (1995). WordNet: A Lexical Database for English.Communications of the ACM Vol. 38, No. 11: 39-41.
5. "WordNets in the World". Global WordNet Association. Retrieved 19 January 2020.
6. S. Baccianella, A. Esuli and F. Sebastiani. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Proceedings of the 7th Conference on Language Resources and Evaluation (LREC'10), Valletta, MT, 2010, pp. 2200–2204.