Computerlinguistik II / Sprachtechnologie

Vorlesung im SS 2018
(M-GSW-10)

Prof. Dr. Udo Hahn

Lehrstuhl für Computerlinguistik
Institut für Germanistische Sprachwissenschaft
Friedrich-Schiller-Universität Jena

http://www.julielab.de

Computational Lexicons, Terminologies & Ontologies

- Computational Lexicons
  - Language-specific information (English, Spanish, German, etc.), cover common-sense knowledge
  - Cover, at best, all linguistic description levels for a lexical item but usually don't
  - Undetermined towards formalization, yet electronically available
- Terminologies
  - Language-independent (though verbally encoded!), cover domain-specific, expert-level knowledge
  - Cover lexico-semantic information only (semantic relations)
  - Informal, computational issues are (usually) of no concern
- Ontologies
  - Language-independent, cover domain-specific, expert-level knowledge
  - cover conceptual information (semantic relations, semantic integrity constraints, rules, etc.)
  - Formal specifications, computational issues are a major concern
  - Formal reasoning: inferences

Examples: Computational Lexicons, Terminologies & Ontologies

- Computational Lexicons
  - WordNet (English) & EuroWordNet
  - GermaNet
  - FrameNet
- (Biomedical) Terminologies
  - Unified Medical Language System (UMLS)
  - BioPortal
  - Open Biological Ontologies (OBO)
    - Gene Ontology (GO)
- Ontologies
  - Formal reasoning (for text understanding)

WordNet

- English WordNet (V3.0)
  - semantic (relation) lexicon of English (general language)
    - no morphology!, no syntax!, no etymology
  - groupings of words into sets of synonyms (synsets)
  - English definitions for lexical entries/synsets (glosses)
  - defines semantic relations between synsets
  - covers (base forms of) nouns, verbs, adjectives, adverbs
  - Size: more than 155,000 lexical entries

http://wordnet.princeton.edu/

WordNet

- EuroWordNet
  - Portuguese, Spanish, Spanish Catalan-Basque
  - French
  - Italian
  - German (licence required), Dutch
  - Russian, Czech, Hungarian, Slovene, …
  - …
- Global WordNet `http://globalwordnet.org`
  - Arabic
  - Mandarin-Chinese
  - Hindi
  - ….

WordNet SynSets and Glosses

- **Nouns**

S: (n) **jump**, **leap** (a sudden and decisive increase) *"a jump in attendance"*
direct hyponym / full hyponym
S: (n) quantum leap, quantum jump (a sudden large increase or advance) *"this may not insure success but it will represent a quantum leap from last summer„*
direct hypernym / inherited hypernym/ sister term derivationally related form
S: (n) **leap**, **jump**, **saltation** (an abrupt transition) *"a successful leap from college to the major leagues"*
S: (n) **jump** ((film) an abrupt transition from one scene to another)
S: (n) **startle**, **jump**, **start** (a sudden involuntary movement) *"he awoke with a start"*
S: (n) **jump**, **parachuting** (descent with a parachute) *"he had done a lot of parachuting in the army"*
S: (n) **jump**, **jumping** (the act of jumping; propelling yourself off the ground) *"he advanced in a series of jumps"; "the jumping was unexpected"*

WordNet Synsets and Glosses

- **Verb**

S: (v) **jump**, **leap**, **bound**, **spring** (move forward by leaps and bounds) *"The horse bounded across the meadow"; "The child leapt across the puddle"; "Can you jump over the fence?"*
S: (v) **startle**, **jump**, **start** (move or jump suddenly, as if in surprise or alarm) *"She startled when I walked into the room"*
S: (v) **jump** (make a sudden physical attack on) *"The muggers jumped the woman in the fur coat"*
S: (v) **jump** (increase suddenly and significantly) *"Prices jumped overnight"*
S: (v) **leap out**, **jump out**, **jump**, **stand out**, **stick out** (be highly noticeable)
S: (v) **jump** (enter eagerly into) *"He jumped into the game"*
S: (v) **rise**, **jump**, **climb up** (rise in rank or status) *"Her new novel jumped high on the bestseller list"*
S: (v) **jump**, **leap**, **jump off** (jump down from an elevated point) *"the parachutist didn't want to jump"; "every year, hundreds of people jump off the Golden Gate bridge"; "the widow leapt into the funeral pyre"*
S: (v) **derail**, **jump** (run off or leave the rails) *"the train derailed because a cow was standing on the tracks"*
S: (v) **chute**, **parachute**, **jump** (jump from an airplane and descend with a parachute)
S: (v) **jump**, **leap** (cause to jump or leap) *"the trainer jumped the tiger through the hoop"*
S: (v) **jumpstart**, **jump-start**, **jump** (start (a car engine whose battery is dead) by connecting it to another car's battery)
S: (v) **jump**, **pass over**, **skip**, **skip over** (bypass) *"He skipped a row in the text and so the sentence was incomprehensible"*
S: (v) **leap**, **jump** (pass abruptly from one state or topic to another) *"leap into fame"; "jump to a conclusion"; "jump from one thing to another"*
S: (v) **alternate**, **jump** (go back and forth; swing back and forth between two states or conditions)

WordNet Relations

- Nouns
  - Hypernyms
    - „Y is a hypernym (more general term) of X, if every X is a (kind of) Y“
  - Hyponyms
    - „Y is a hyponym (more specific term) of X, if every Y is a (kind of) X“
    - Y is hyponym of X ↔ X is a hypernym of Y
  - Coordinate terms
    - „Y is a coordinate term of X, if X and Y share a hypernym“
  - Holonyms
    - „Y is a holonym (whole) of X, if (every/some?) X is a part of Y“
  - Meronyms
    - „Y is a meronym (part) of X, if (every/some?) Y is a part of X“
    - Y is a meronym of X ↔ X is a holonym of Y

WordNet Relations

- Verbs
  - Hypernyms
    - „the verb Y is a hypernym (more general term) of the verb X, if the activity X is a (kind of) Y“
      - e.g., *travel* to *movement*
  - Troponyms
    - „the verb Y is a troponym of the verb X, if the activity Y is doing X in some manner“
      - e.g., *lisp* to *talk*
  - Entailment
    - „the verb Y is entailed by the verb X, if by doing X you must be doing Y“
      - e.g., *snoring* by *sleeping*
  - Coordinate terms
    - „Y is a coordinate verb of X, if X and Y share a hypernym“

WordNet Relations

- Adjectives
  - Related nouns
  - Participle of verb
- Adverbs
  - Root adjectives

WordNet V3.0 Statistics

POS	Unique Strings	SynSets	Word-Sense Pairs (word - #synset pairs)
Noun	117,100	82,100	146,300
Verb	11,500	13,800	25,000
Adj	21,500	18,200	30,000
Adv	4,500	3,600	5,600
Σ	155,300	117,700	206,900

http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html

WordNet V3.0 Statistics

POS	Monosemous Words / senses	Polysemous Words	Polysemous senses
Noun	101,900	16,000	44,400
Verb	6,300	5,300	18,800
Adj	16,500	5,000	14,400
Adv	3,700	700	1,800
Σ	128,400	27,900	79,500

http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html

WordNet V3.0 Statistics

POS	Average polysemy* (incl. monosemous words)	Average polysemy* (excl. monosemous words)
Noun	1.24	2.79
Verb	2.17	3.57
Adj	1.40	2.71
Adv	1.25	2.50

* Number of synsets that contain the word

http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html

GermaNet

- **Lexikalische Einheiten sind nach Lesarten gruppiert:**

- "Bank" has 2 readings
  - Reading 1: *[**Bank**, {Sitzbank}]* *(bench)*
  - Reading 2: *[**Bank**, {Geldinstitut}]* *(financial institution)*

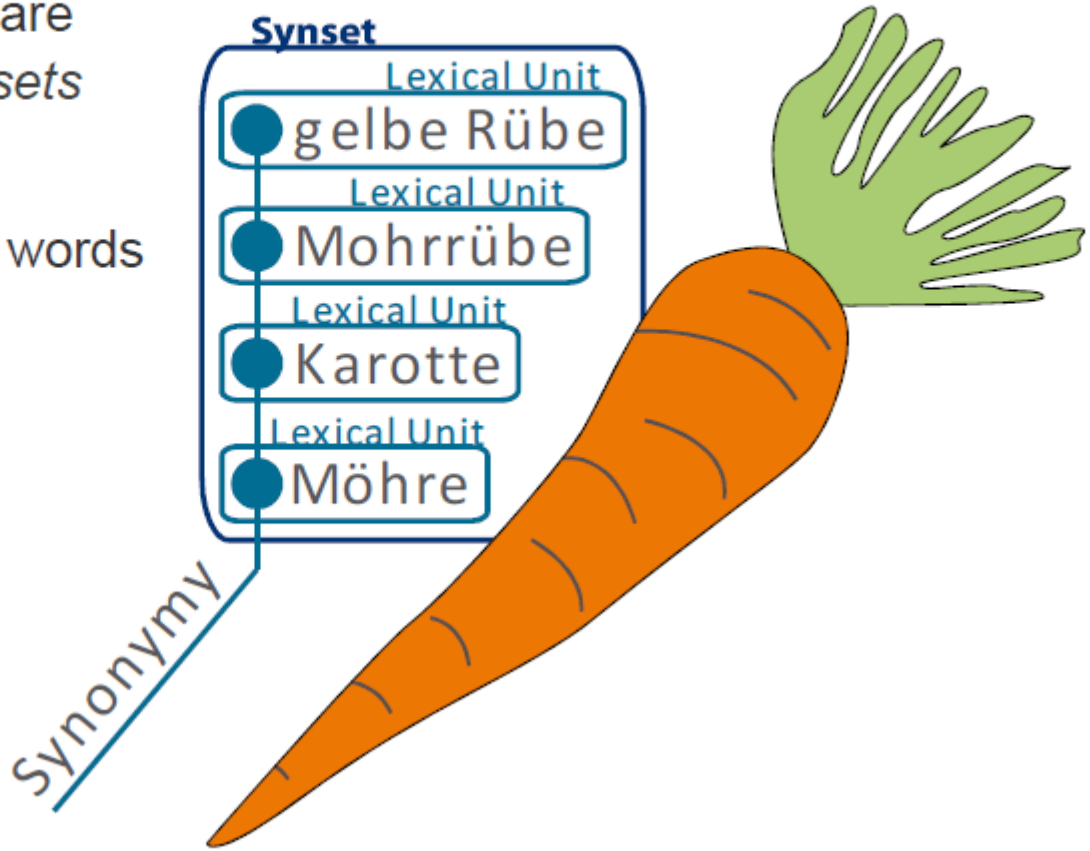
- "Leiter" has 3 readings
  - Reading 1: *[**Leiter**, {Steiggerät}]* *(ladder)*
  - Reading 2: *[**Leiter**, {Verantwortlicher, Anführer}]* *(leader)*
  - Reading 3: *[**Leiter**, {stromleitender Stoff}]* *(electric conductor)*

Quelle: Henrich & Horber, 2015

GermaNet: SynSets

Synsets in GermaNet

- Semantic concepts are represented by *synsets*
- A synset is a set of (near-)synonymous words

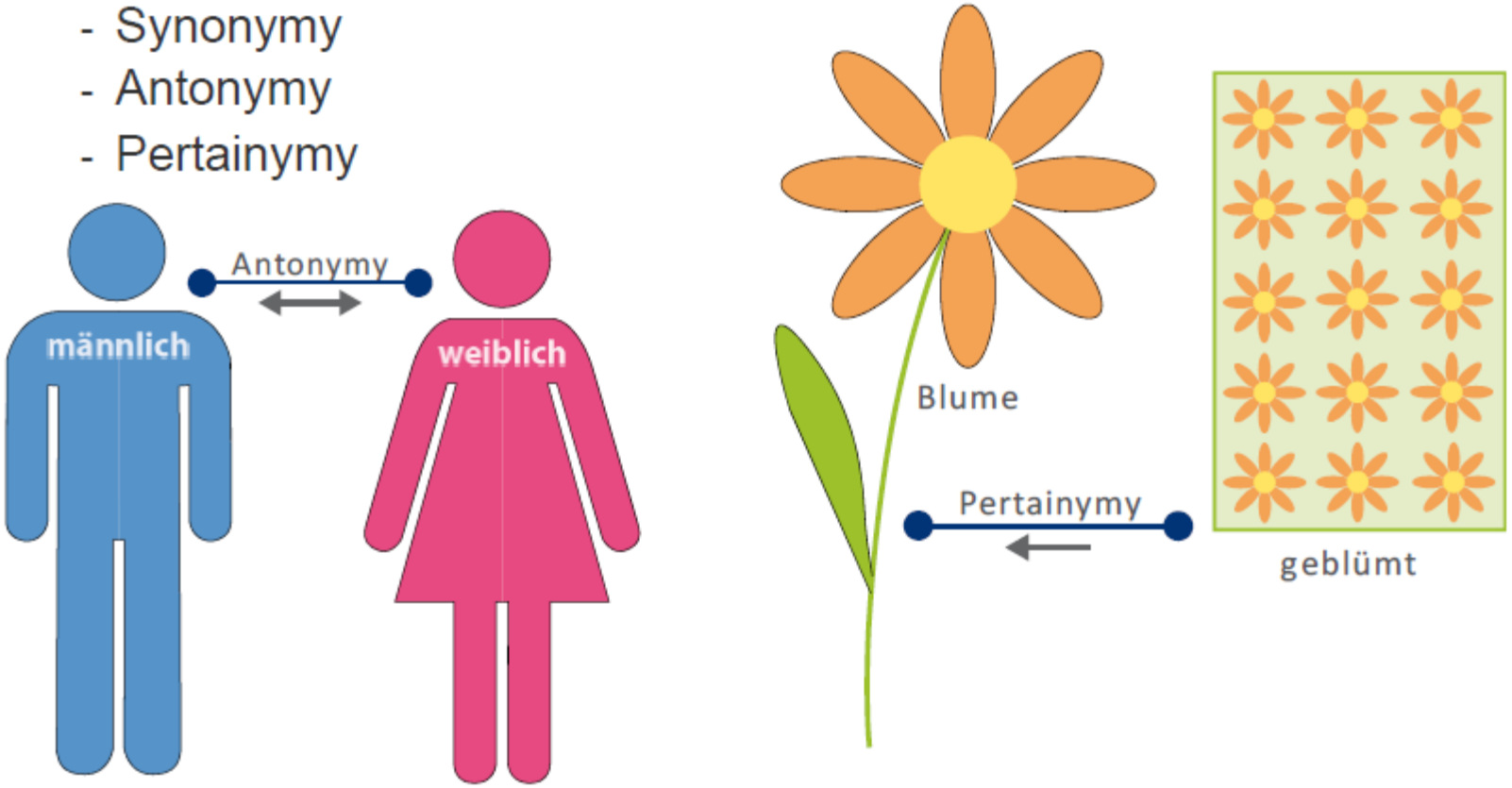


Quelle: Henrich & Horber, 2015

GermaNet: Lexical Relations

Lexical Relations in GermaNet

- Lexical relations hold between two lexical units
  - Synonymy
  - Antonymy
  - Pertainymy

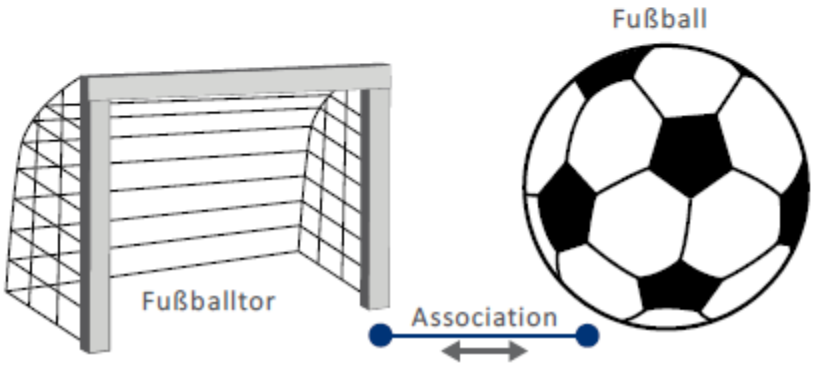


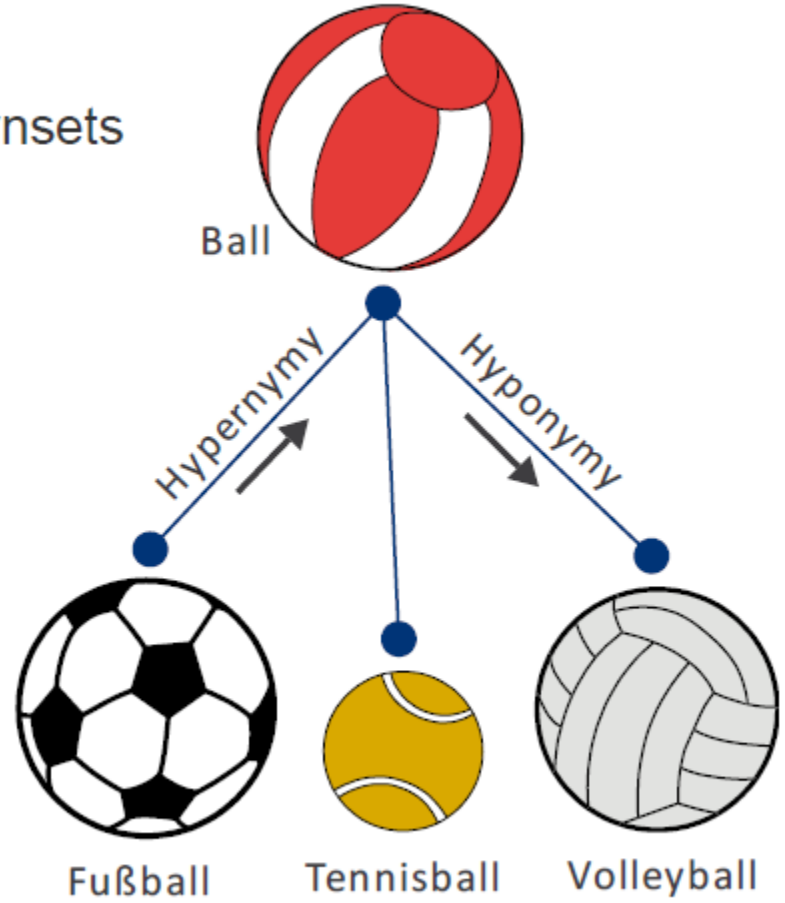
Quelle: Henrich & Horber, 2015

GermaNet: Conceptual Relations

Conceptual Relations in GermaNet

- Conceptual relations hold between synsets
  - Hypernymy and (co-)hyponymy
  - Part-whole relations
  - Entailment
  - Causation
  - Association



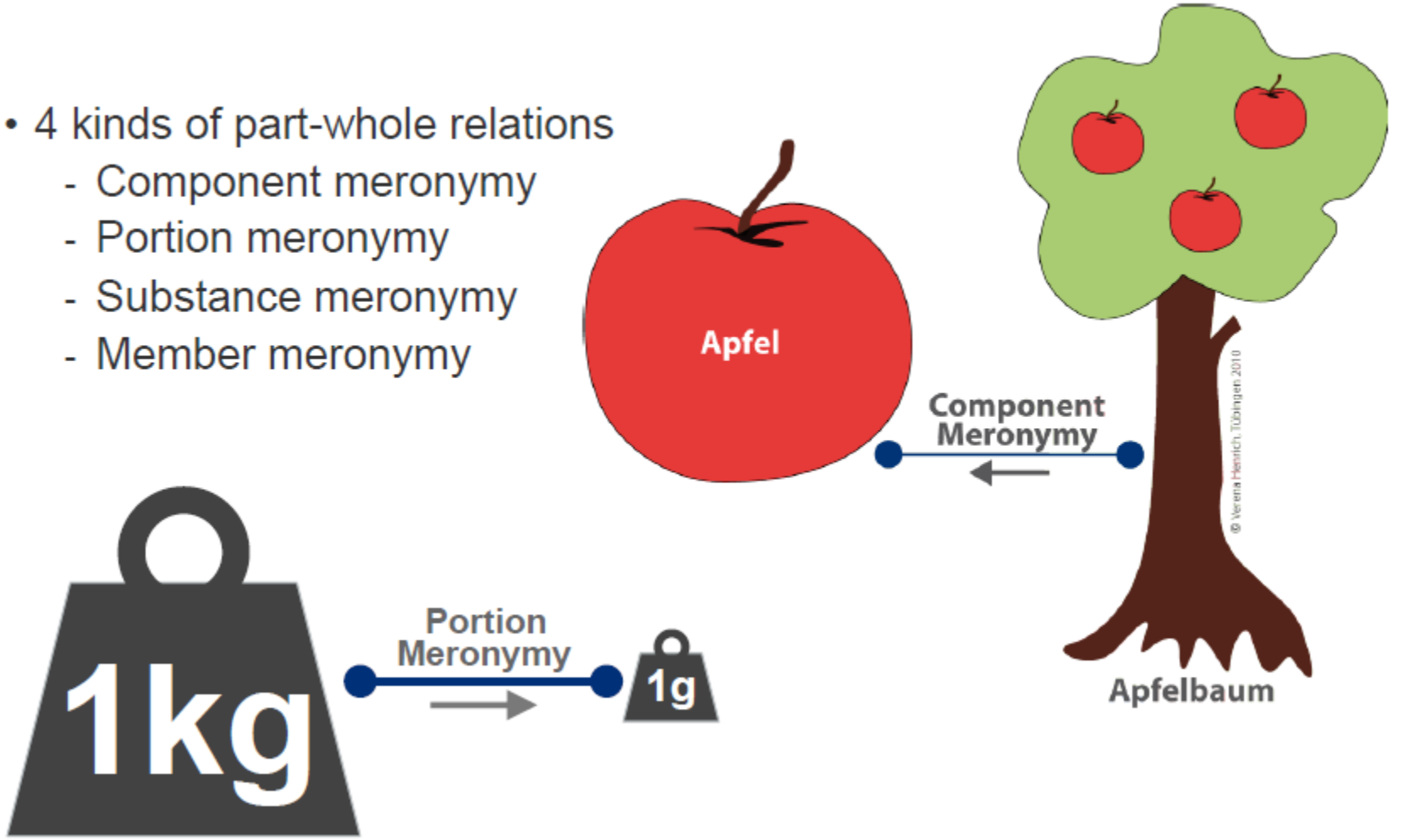


Quelle: Henrich & Horber, 2015

GermaNet: Conceptual Relations

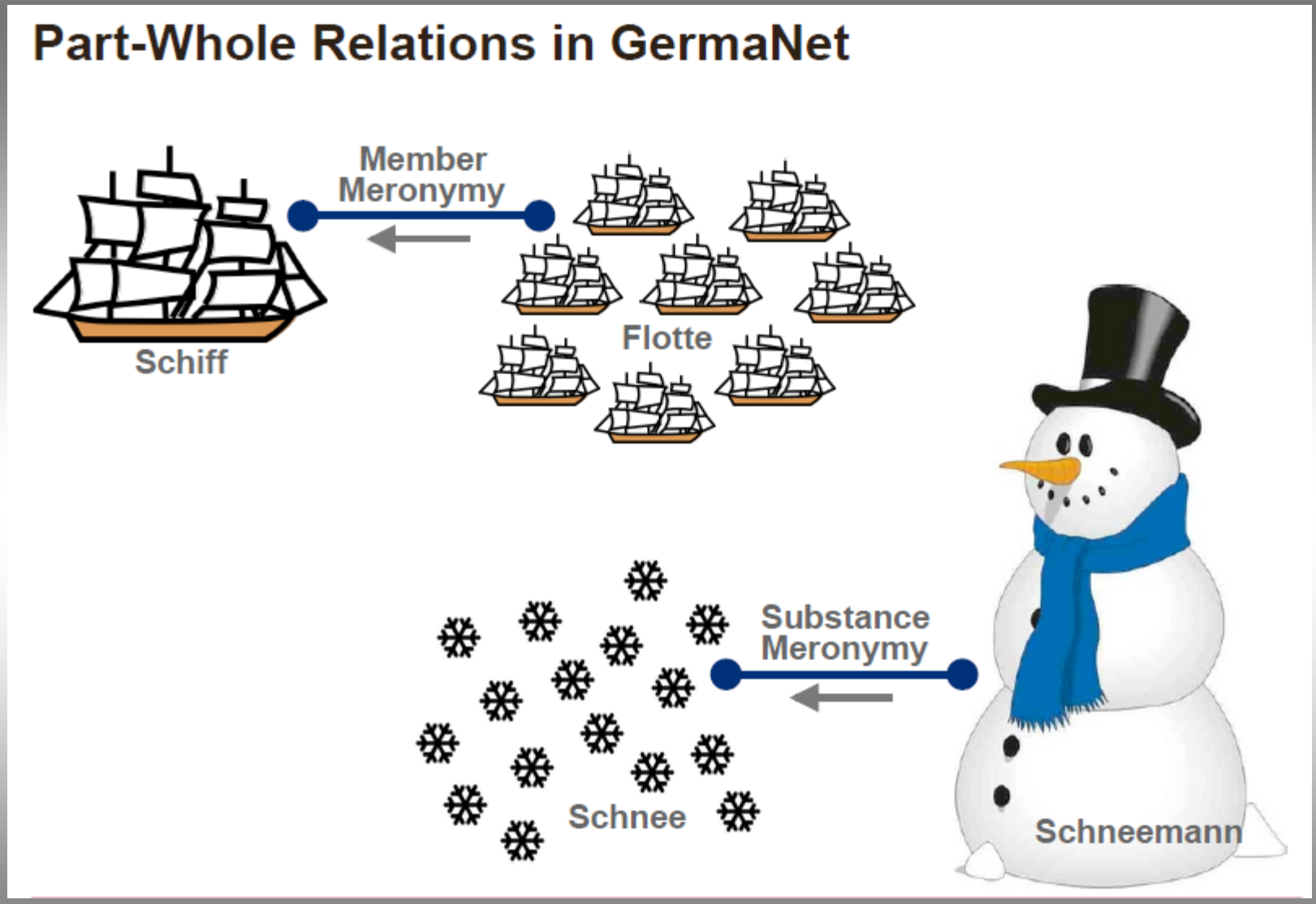
Part-Whole Relations in GermaNet

- 4 kinds of part-whole relations
  - Component meronymy
  - Portion meronymy
  - Substance meronymy
  - Member meronymy



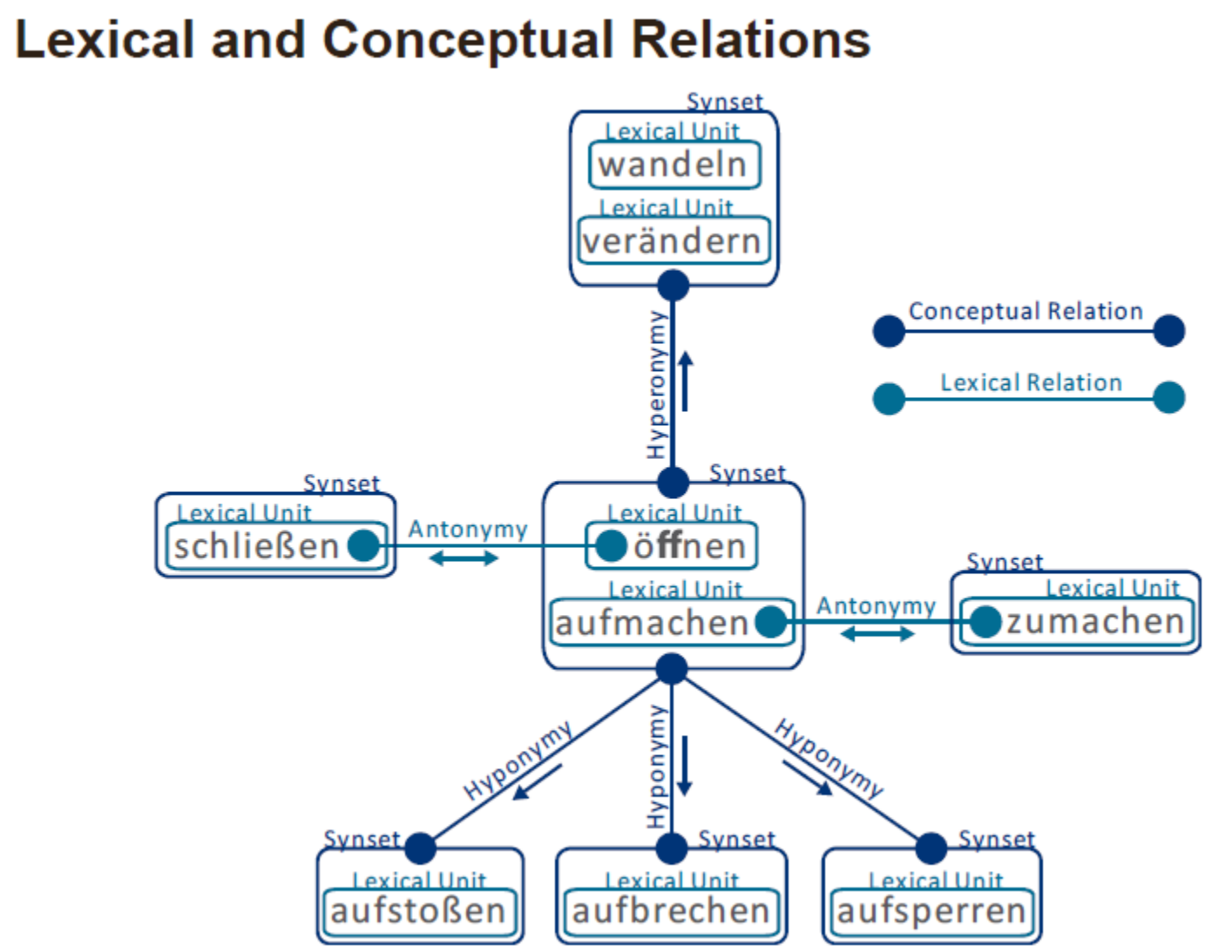
Quelle: Henrich & Horber, 2015

GermaNet: Conceptual Relations



Quelle: Henrich & Horber, 2015

GermaNet: Relations



Quelle: Henrich & Horber, 2015

GermaNet: semantische Felder

38 Semantic Fields

- Allgemein ‘general’
- Artefakt ‘artifact’
- Attribut ‘attribute’
- Besitz ‘possession’
- Bewegung ‘motion’
- Form ‘shape’
- Gefühl ‘feeling’/‘emotion’
- Geist ‘spirit’
- Geschehen ‘event’
- Gesellschaft ‘social’
- Gruppe ‘group’
- Kognition ‘cognition’
- Kommunikation ‘communication’
- Konkurrenz ‘competition’
- Kontakt ‘contact’
- Körper ‘body’
- Körperfunktion ‘bodily function’
- Lokation ‘location’
- Menge ‘quantity’
- Mensch ‘person’
- Motiv ‘motive’
- Nahrung ‘food’
- Naturgegenstand ‘natural object’
- Naturphänomen ‘natural phenomenon’
- Ort ‘place’
- Pertonym ‘pertainym’
- Perzeption ‘perception’
- Pflanze ‘plant’
- privativ ‘privative’
- Relation ‘relation’
- Schöpfung ‘creation’
- Substanz ‘substance’
- Tier ‘animal’
- Tops ‘tops’
- Veränderung ‘change’
- Verbrauch ‘consumption’
- Verhalten ‘behavior’
- Zeit ‘time’

Quelle: Henrich & Horber, 2015

GermaNet : Verb Frames

- Verbbedeutung wird durch Verb-Frames ergänzt
- Verb-Frames beschreiben syntaktische Strukturen wie Valenzen, Subkategorisierungen u.ä. (Bezug auf CELEX Lexikonformat)
- Interlingualer Index (ILI) zur Anbindung an das englische WordNet
- Verlinkung geschieht über lexikalische und konzeptuelle Relationen (derzeit ca. 29.000 Links)

http://www.sfs.uni-tuebingen.de/GermaNet/

GermaNet

Readings for "unterhalten"

1. (v, Besitz) [unterhalten, pflegen] – *über etwas verfügen*
   - **Ex./frames:** *Sie unterhalten gute Beziehungen zu ihren Nachbarn. (NN.AN.Pp)*
   - **Synonyms:** [pflegen]
   - **Hypernyms:** [haben, besitzen]

2. (v, Gefühl) [unterhalten] – *sich auf angenehme Weise die Zeit vertreiben*
   - **Ex./frames:** *Sie hat sich blendend unterhalten. (NN.AR.BM)*
   - **Wiktionary paraphrase:** reflexiv: sich amüsieren, zerstreuen
   - **Hypernyms:** [vergnügen]
   - **ILI synonym:** [amuse] occupy in an agreeable, entertaining or pleasant fashion

3. (v, Gefühl) [unterhalten] – *für Zerstreuung/Zeitvertreib sorgen*
   - **Ex./frames:** *Er unterhielt seine Gäste mit Musik. (NN.AN.Bs)*
   - **Wiktionary paraphrase:** etwas aufsagen, vorführen, vorspielen zur Zerstreuung
   - **Hypernyms:** [vergnügen, amüsieren]
   - **ILI synonym:** [entertain] provide entertainment for

4. (v, Gesellschaft) [unterhalten] – *etw. halten/einrichten/ betreiben und dafür aufkommen*
   - **Examples/frames:** *Er unterhält einen Reitstall. (NN.AN)*
   - **Hypernyms:** [führen]
   - **Hyponyms:** [bewirtschaften]
   - **ILI synonym:** [keep] supply with necessities and support

5. (v, Kommunikation) [unterhalten] – *ein Gespräch führen*
   - **Examples/frames:** *Er unterhielt sich nur über die Arbeit. (NN.AR.Pp); Er unterhielt sich nur mit mir. (NN.AR.Bo)*
   - **Wiktionary paraphrase:** reflexiv: mit jemandem reden, erzählen
   - **Hypernyms:** [austauschen]
   - **Hyponyms:** [klönen] [labern] [palavern] [philosophieren] [plauschen] [plaudern, schwatzen, schnattern]
   - **ILI synonym:** [speak] exchange thoughts; talk with

6. (v, Verbrauch) [unterhalten, alimentieren] – *für jmds. Lebensunterhalt aufkommen*
   - **Ex. /frames:** *Er unterhält eine acht-köpfige Familie. (NN.AN)*
   - **Wiktionary paraphrase:** Unterhalt zahlen, jm. unterstützen
   - **Synonyms:** [alimentieren]
   - **Hypernyms:** [ernähren, nähren]
   - **ILI hypernyms:** [feed] give food to; [support] support materially or financially

Quelle: Henrich & Horber, 2015

GermaNet-Struktureigenschaften

- WordNets sind hierarchisch über die Hyponymie / Hypernymie-Relation und partonomisch über diverse Teil-Ganzes-Relationen strukturiert
- Im Unterschied zum englischen WordNet gibt es nicht mehrere Hierarchien – GermaNet ein vollständig verbundener Graph mit einem gemeinsamen (künstlichen) Wurzelknoten (GNROOT), der alle SynSets subsumiert
- Künstlich eingeführte Knoten beschreiben nicht-lexikalisierte Konzepte
  - Vermeidet unnötige Ko-Hyponymien
- GermaNet erlaubt Kreuzklassifikation (ein SynSet kann mehrere Hypernyme haben)

http://www.sfs.uni-tuebingen.de/GermaNet/

GermaNet (V 9.0 – April 2014)

- **121.810 Lexeme**
  - 93.630 Nomen
  - 14.330 Verben
  - 13.8500 Adjektive
- **93.250 Synsets**
  - 71.500 Nomen
  - 11.030 Verben
  - 10.650 Adjektive
- **105.900 Relationen**
  - 95.000 Hypernym/Hyponym
  - 10.000 Holonym/Partonym

http://www.sfs.uni-tuebingen.de/GermaNet/

GermaNet (V 12.0 – May 2017)

- 154.814 Lexeme

- 120.032 Synsets

- 133.652 konzeptuelle Relationen
- 4.210 lexikalische Relationen (ohne Synonymie)

- Größe des ILI: 28.567
- Anzahl Wiktionary-Beschreibungen: 29.552

http://www.sfs.uni-tuebingen.de/GermaNet/

FrameNet

- English FrameNet
  - semantic frames of English (script-style)
    - no morphology!, no syntax!, no etymology
  - English-style, semi-formal definitions for lexical entries
  - Statistics (Version 1.3)
    - 11,000 lexical units
    - 1,050 semantic frames
    - 135,000 example sentences for frames (taken from the British National Corpus [BNC] and US newswire)

http://framenet.icsi.berkeley.edu/

http://framenet.icsi.berkeley.edu/index.php?option=com_content&task=view&id=17881&Itemid=66/

- Try out: FrameGrapher

FrameNet Entry

FrameNet Data Search for jump

Lexical unit search results: Closest match is jump...

Lexical Unit	Frame
jump.v	Self_motion
jump.v	Traversing
jump.v	Change_position_on_a_scale
jump.v	Attack
jumper.n	Clothing
jumping.a	Lively_place
jumpsuit.n	Clothing

FrameNet Entry (cont.)

Self_motion

Definition: The Self_mover, a living being, moves under its own power in a directed fashion, i.e. along what could be described as a Path, with no separate vehicle.

FEs: **Core:**

Area [Area] **Semantic Type** Location

Area is used for expressions which describe a general area in which motion takes place when the motion is understood to be irregular and not to consist of a single linear path. Note that this FE should not be used for cases when the same phrase could be used with the same meaning with a non-motion target, since these should be annotated with the Place FE.

Direction [dir]

The direction that the Self_mover heads in during the motion.

Goal [Goal] **Semantic Type** Goal

Goal is used for any expression which tells where the Self_mover ends up as a result of the motion.

Path [Path] **Semantic Type** Path

Path is used for any description of a trajectory of motion which is neither a Source nor a Goal. This includes "middle of path" expressions.

Self_mover [SMov] **Semantic Type** Sentient

Self_mover is the living being which moves under its own power. Normally it is expressed as an external argument.

Source [Src] **Semantic Type** Source

Source is used for any expression which implies a definite starting-point of motion. In prepositional phrases, the prepositional object expresses the starting point of motion. With particles, the starting point of motion is understood from context.

Wiktionary

- **Multilinguales Wörterbuch**
  - **derzeit 230 Sprachen**
- **625.973 deutschsprachige Einträge**
  - **Struktur s. Beispiel „Hebamme“**
- **Online:**
  - **https://de.wiktionary.org/wiki/Wiktionary:Hauptseite**

Wiktionary-Eintrag „Hebamme“ (1/3)

Hebamme (Deutsch) [Bearbeiten]

Substantiv, *f* [Bearbeiten]

Worttrennung:
Heb·am·me, Plural: Heb·am·men

Aussprache:
IPA: ['he:p,ʔamə], ['he:,bamə]
Hörbeispiele: Hebamme (Info), —

	Singular	Plural
Nominativ	die Hebamme	die Hebammen
Genitiv	der Hebamme	der Hebammen
Dativ	der Hebamme	den Hebammen
Akkusativ	die Hebamme	die Hebammen

Bedeutungen:
[1] früher kundige, heutzutage an einer speziellen Lehranstalt ausgebildete und staatlich geprüfte Frau, die während der Schwangerschaft, bei der Geburt, während des Wochenbetts und der Stillzeit die werdende beziehungsweise frischgebackene Mutter berät, betreut und ihr Hilfe leistet

- *übertragen umgangssprachlich:*

[2] Gerät, mit dem Korken aus einer Flasche gezogen werden
[3] Gerät, mit dem sich Flaschen mit Kronenkorken-Verschluss öffnen lassen
[4] Assistent eines Regisseurs

[1] das Foto zeigt die ungarische *Hebamme* und Ärztin Ágnes Geréb beim Kontrollieren der Herztöne des Fötus während der Wehen bei einer Hausgeburt in Budapest; Aufnahme vom 25. Februar 2001

Herkunft:
[1] Bei dem auf das deutsche Sprachgebiet beschränkten Ausdruck[1] handelt es sich um ein seit dem 9. Jahrhundert[2][3] bezeugtes Erbwort, dessen althochdeutsche Formen *heviana* → *goh*[2], *hevianna* → *goh*[1][2][3], *hevanna* → *goh*[1][2][3] sowie *hevamma* → *goh*[2][3] lauteten. Diese leben mundartlich heute noch fort (*hebane* und so weiter).[2] Die Herkunft des Zweitglieds *-anna* ist dunkel.[2] Vermutlich sind die Formen einerseits aus dem unter *heben* dargestellten Verbalstamm[3], der sich offenbar auf das Hochheben des Kindes unmittelbar nach der Geburt bezieht[2], und althochdeutschem *ana* → *goh* ‚Großmutter‘ (siehe *Ahne*) zusammengesetzt, wodurch die Komposita also eigentlich eine ‚Großmutter beziehungsweise alte Frau, die das Neugeborene aufhebt‘ bezeichnen[3]. Andererseits könnte es sich vermutlich auch ursprünglich um eine Bildung mit dem althochdeutschen Suffix *-ina* zu dem althochdeutschen Verb *heffen* → *goh* ‚heben‘ als ‚die Heberin‘ nach dem Muster von althochdeutsch *meisterina* → *goh* ‚Leiterin‘ handeln.[1]

Wiktionary-Eintrag „Hebamme“ (2/3)

Die Umdeutung des Grundwortes *-ana* über verstärkendes *-anna* zu *-amma* ‚Amme‘ vollzieht sich bereits im Althochdeutschen und setzt sich im Mittelhochdeutschen fort[3], in dem Formen wie *hebeamme* → *gmh*[1][2][3], *hebamme* → *gmh*[1] und *hefamme* → *gmh*[3] neben gleichbedeutend *hebemuoter* → *gmh*[3] bezeugt sind. Im Neuhochdeutschen setzt sich »Hebamme« gegen landschaftliche Ausdrücke wie *Bademutter, Hebemutter, Kindermutter, weise Frau* und *Wehmutter* (eine Prägung Luthers) durch und wird zur allgemein gültigen Berufsbezeichnung.[3]

[2] Die ab 1960 bezeugte und von der Kellnersprache in die Umgangssprache übernommene Bedeutungsübertragung fußt darauf, dass der Vorgang im Prinzip einer (Zangen-)Geburt ähnelt und außerdem eine Anspielung auf *heben* ‚trinken‘ vorliegt.[4][5]

[3] Diese Bedeutung ist ebenfalls ab 1960 bezeugt und wurde aus der Kellnersprache in die Umgangssprache übernommen.[4][5]

[4] Diese Bedeutung ist aus der Filmsprache in die Umgangssprache eingegangen und dort ab 1920 bezeugt.[4] Das umgangssprachliche Übertragungsmotiv fußt darauf, dass der Regieassistent eine Art Geburtshelfer bei der Filmaufnahme ist.[4]

Synonyme:

[1] Geburtshelferin

[1] *fachsprachlich (Medizin):* Obstetrix

[1] *veraltet:* Bademuhme, Bademutter, Frodfrau, Hebemutter, Kindermutter, Kindmuhme, Küchleinmutter, Sage Femme, Wehfrau, Wehmutter, weise Frau/Weisfrau/Weisefrau, Weisemutter

[1] *umgangssprachlich*[6]*:* Empfangsdame, Grapsche, Mutter Graps, Storch/*zumeist:* Frau Storch

[1] *Brasilien (Rio Grande do Sul, Santa Catarina):* Parteira

[1] *Nordamerika (Pennsylvaniadeutsch):* Wartefrau

[2] Korkenzieher

[2] *bundesdeutsch landschaftlich:* Korkzieher, Pfropfenzieher

[2] *österreichisch:* Stoppelzieher

[2] *schweizerisch:* Zapfenzieher

[3] Flaschenöffner

[4] Regieassistent

Wiktionary-Eintrag „Hebamme“ (3/3)

Sinnverwandte Wörter:

[1] Doula

Gegenwörter:

[1] Engelmacherin

Männliche Wortformen:

[1] *Deutschland, Schweiz offiziell:* Entbindungspfleger

[1] *Österreich offiziell:* Hebamme

[1] *umgangssprachlich:* Hebammer[4], Hebammerich[4]

Unterbegriffe:

[1] Beleghebamme, Familienhebamme, Lehrhebamme, Privathebamme

Beispiele:

[1] Die *Hebamme* half der gebärenden Schwangeren bei der Niederkunft und der Entbindung.

[1] „«Eine Aufgabe für *Hebammen»*, erwiderte der Arzt mürrisch. «Dazu bin ich hergebeten worden?»“[7]

[1] „Die Erstgebärende ist bei den *Hebammen* am wenigsten beliebt, denn sie ist unerfahren, und nicht immer wird sie behutsam behandelt.“[8]

[1] „„Hole so schnell wie möglich den Sanitätsrat Querfot und die *Hebamme* Kakeldütt.[…]‘“[9]

[1] „Sie gratulieren mir, und ich kann nicht umhin, diese Gratulation in Gedanken weiterzugeben an eine längst verstorbene Frau, eine gewisse Josefine H., die in meinem Geburtsschein als *Hebamme* eingetragen ist.“[10]

[1] „Das komplexe Wechselspiel zwischen der Anatomie des aufrechten Ganges und den Anforderungen eines großen Gehirns ist dennoch faszinierender als verklärte Romanerzählungen. Und entgegen anders lautenden Behauptungen ist vielleicht der Beruf der *Hebamme* das älteste Gewerbe der Welt.[11]

[1] „Lea entband in der Lindenallee, mit Hilfe Doktor Aarons und einer *Hebamme*.“[12]

[1] „Die *Hebamme* hieß Frau Rassmann und wohnte in Mathesdorf, fünf Straßenbahnstationen vom Haus der Eckerts entfernt.“[13]

[2] „Für alte Weine nutze ich manchmal auch die sogenannte *Hebamme,* eine Art Zange, die man zwischen Flaschenhals und Korken schiebt.“[14]

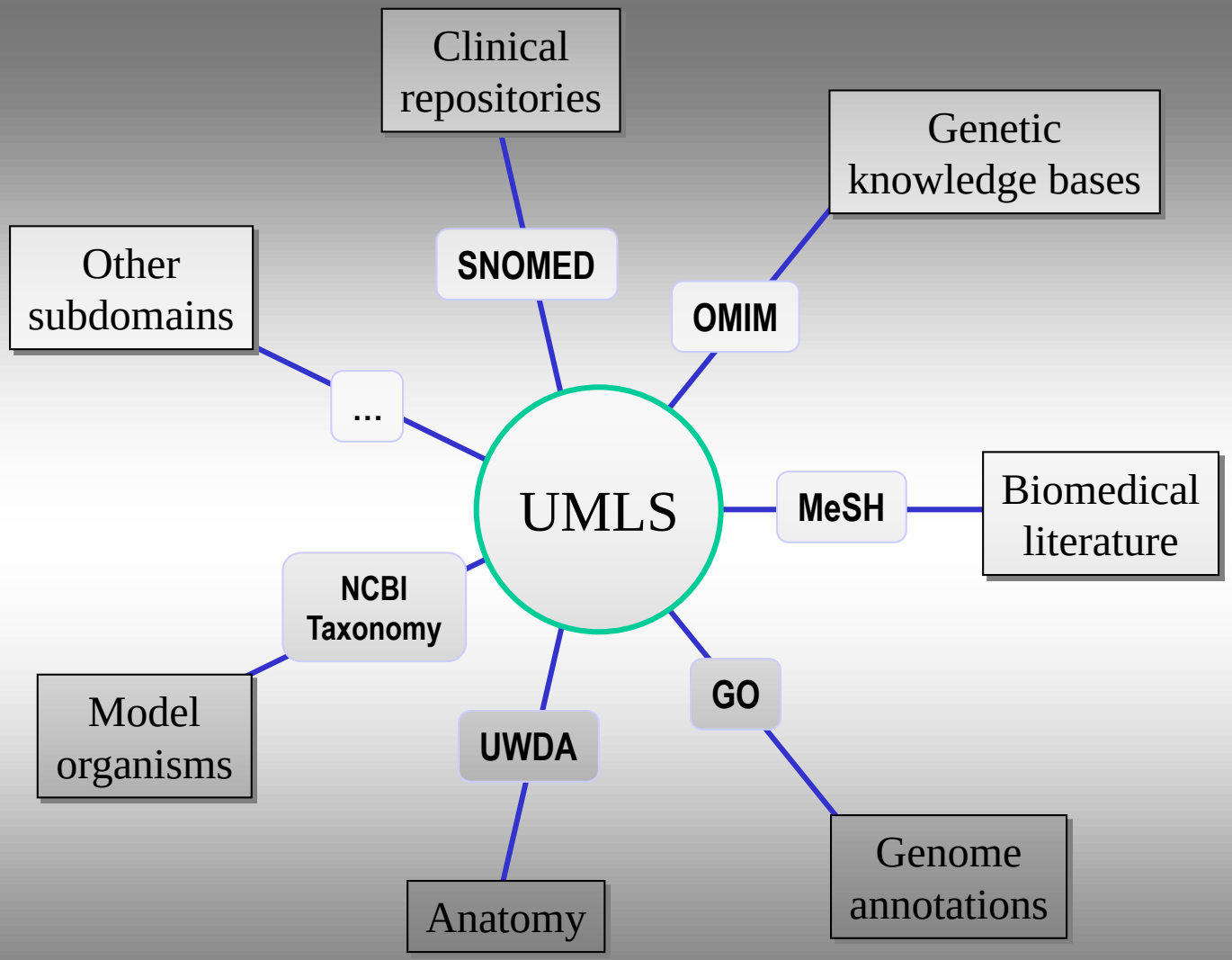
(Biomedical) Terminologies

- Sublanguages: domain-specific
- Relational Encoding
  - Is-a
  - Part-of

UMLS – Unified Medical Language System

- https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/statistics.html
- https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/MSH/stats.html
- Purpose: clinical coding, billing, document retrieval, …
  - Umbrella system covering more than 153 terminologies
  - Size: 14,000,000 terms (10M English); <3.700.000 concepts, >>13,000,000 relations
  - Content: (almost) the whole of (clinical) medicine
  - Lexical semantics: thesaurus relations for taxonomies, partonomies, also other light-weight semantics (approximately 80 additional relation types)
  - Basic and variant word forms, and (quite complex) NPs
  - (English) Specialist Lexicon uses conceptual grounding of UMLS for NLP applications

UMLS Thesauri



UMLS Tables

UMLS Unified Medical Language System

NLM NATIONAL LIBRARY OF MEDICINE

Concept 1	*relation*	*Concept 2*
RIGHT-SIDE-OF-HEART	narrower_rel	HEART
LEFT-SIDE-OF-HEART	part_of	HEART
ANGINA-PECTORIS	has_location	HEART
HEART	has_part	HEART-ATRIUM
HEART	has_part	MITRAL-VALVE
WALL-OF-HEART	part_of	HEART
BRONCHIAL-TUBERCULOSIS	has_location	BRONCHI
BRONCHIAL-TUBERCULOSIS	narrower_rel	TUBERCULOSIS
SARCOMA	sibling	CARCINOMA
LENS-CRYSTALLINE	part_of	EYE
ACUTE-MYELOID-LEUKEMIA	has_location	BONE-MARROW
RIGHT-HAND	is_a	HAND
ALLERGIC-REACTION	associated_with	DERMATITIS-ATOPIC
LUNG	broader_rel	ATELECTASIS

anatomical concepts
pathological concepts

BioPortal

(June 2018)

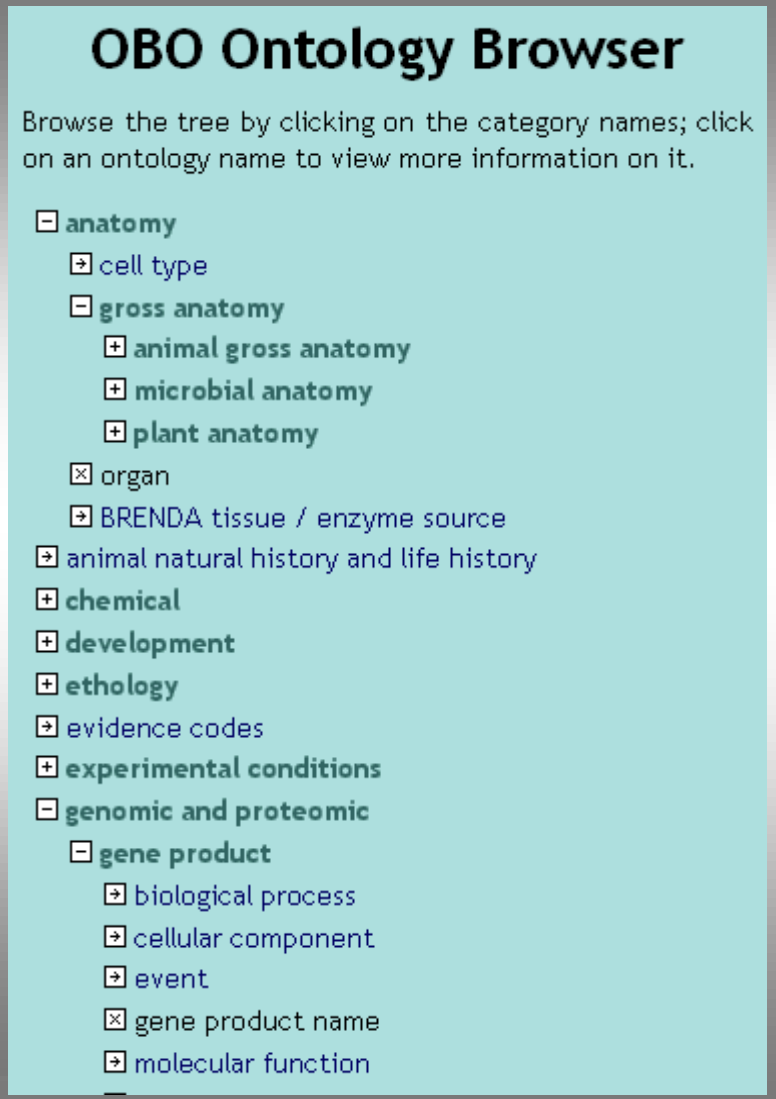
- Hosted by the U.S. National Center for Biomedical Ontology
- 716 biomedical ontologies
- 9.180.000 classes
- 95,500,000 annotations
- Searching all ontologies
- Recommending most suitable ontologies (given a textually provided interest profile)

- http://bioportal.bioontology.org/

Open Biological Ontologies (OBO)

http://obo.sourceforge.net

- Coverage:
  - Anatomy (cells, human, model organisms, etc.)
  - Chemical entities
  - Experimental conditions
  - Genomics, proteomics
  - …
- Structured controlled vocabularies (thesauri)
- Basic Relations: `is-a`, `part-of`
- OBO entry: ID, concept name, textual definition, synonyms



OBO Statistics

(June 2018)

- More than 209 OBO ontologies
  - 5.382000 terms, 19.000 properties, 479000 individuals
  - http://www.ebi.ac.uk/ols/index
- about 50% of them contain more than 1000 terms:
  - 2 x > 25 000 terms: NCI Thesaurus, FMA (Human Anatomy), Gene Ontology (GO), ChEBI (chemicals)
  - 5 x 10 000-25 000: disease ontology, MeSH "ontology", mouse anatomy stages
  - 18 x 1000 -10 000 terms: molecule role (chemicals, protein by function), human, mouse, fly, fish anatomy (some: developmental anatomy), etc.
- Less than 1000 terms: cell ontology, pathway ontology, MGED (Microarray Gene Expression Database), relationship ontology (amongst others)
- Rapidly growing! – check out every day (o.k., week is also fine)

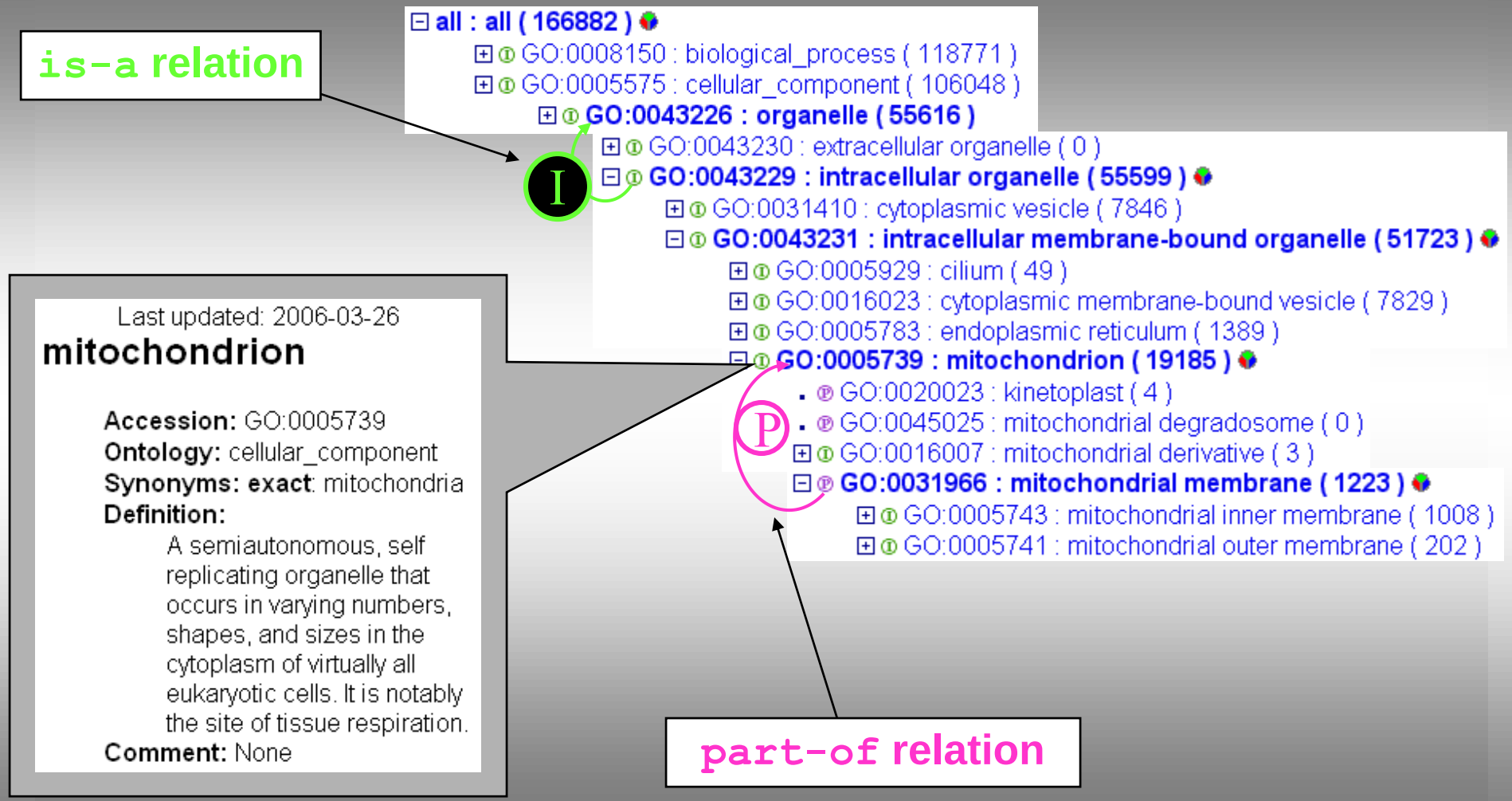
  `http://www.obofoundry.org`

Gene Ontology (GO) *June 2016*

- Purpose: Data annotation and integration for genes and gene products (cross-species)
- **Coverage: Three ontologies in one for molecular biology**
  - **cellular component**: location of a gene product, within (sub)cellular structures and macromolecular complexes, e.g., *nucleus* or *ribosome*
  - **molecular function**: the tasks performed by individual gene products at the biochemical level, e.g., *enzyme* or *transporter*
  - **biological process**: biological goals to which a gene product contributes; that process is accomplished by ordered assemblies of molecular functions, e.g., *mitosis* or *cell growth*
- 44,200 classes (95,6% w./ verbal definitions)
- **2 base relations; 60,500 relation instances**
  - Specific/general (88%) (*mitotic chromosome* `is-a` *chromosome*)
  - Part/whole (12%) (*telomere* `part-of` *chromosome*)

`http://www.geneontology.org`

Snapshot of *GO*



General Shortcomings

- Category descriptions, at best, are verbally defined
- Relations are usually undefined, their names appeal to human/expert intuition
- (Almost) No attempt at interoperability
- Lots of unlinked fragments (still a long way to go to some sort of ‘Bio-UMLS’)

Ontologies

- Formal Reasoning
- Conceptual Computation

Why Conceptualize?

- Nomenclatures, thesauri, ontologies, …
- "Mapping problem" due to term variation
  - Natural language ⇨ domain knowledge

"Mapping Problem" (1/2)

Problem: Mapping a textual occurrence of a bio entity (text token, term) to its ontological category (type)

- Orthographic variations
  - Hyphens, slashes, spaces (e.g., NF-KB, NF KB, NF/KB, NFKB)
  - Upper/lower cases (e.g., NF-KB, NF-kb)
  - Spelling variations (e.g., tomour vs. tumor, oestrogen vs. estrogen, alpha vs. α)
- Lexical and phrasal variations
  - Acronyms (e.g., RAR vs. retinoic acid receptor)
  - Different reductions (e.g., SB2 gene vs. SB2, thyroid hormone receptor vs. thyroid receptor)

"Mapping Problem" (2/2)

- Semantic variations (n:m token-type relations)
  - n:1 Synonyms (e.g., in FlyBase: EST-6 vs. Esterase 6 vs. carboxyl ester hydrolase)
  - 1:m Ambiguity as polysemy (e.g., 'per' in FlyBase: period gene vs. clock gene)

Why Is Bio Terminology So Hard?

V-SNARE

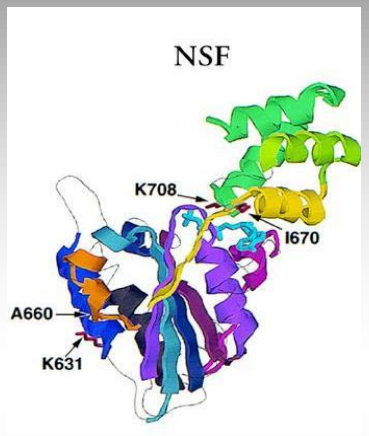
Vesicle SNARE

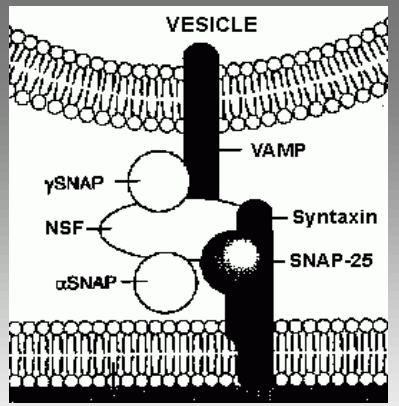
SNAP Receptor

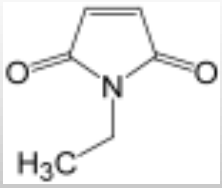
Soluble NSF Attachment Protein

N-Ethylmaleimide-Sensitive Fusion Protein

Maleic acid N-ethylimide





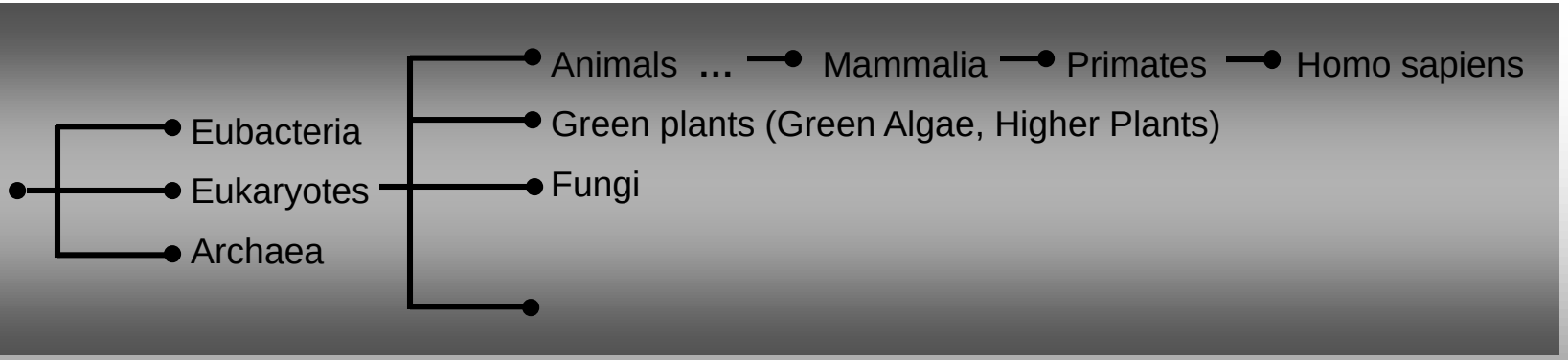


Vesicle Soluble Maleic acid N-ethylimide Sensitive Fusion Protein Attachment Protein Receptor

Why Conceptualize?

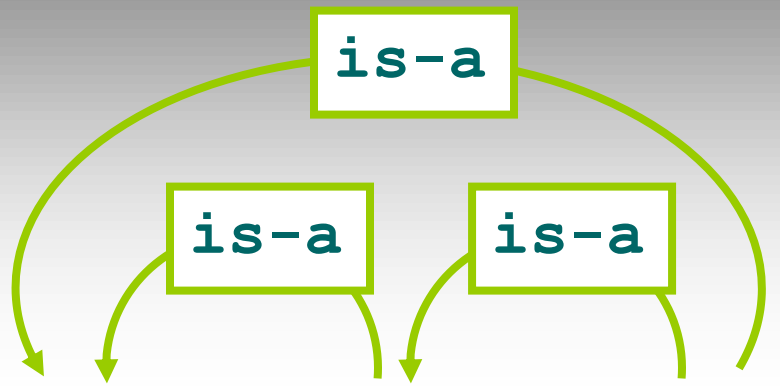
- Nomenclatures, thesauri, ontologies, …
- "Mapping problem" due to term variation
  - Natural language ⇨ domain knowledge
- **"Structure computing" on knowledge structures**
  - **Lexical look-up**
  - **Relational navigation (general-specific, is-a)**
  - **Formal reasoning (inferencing)**

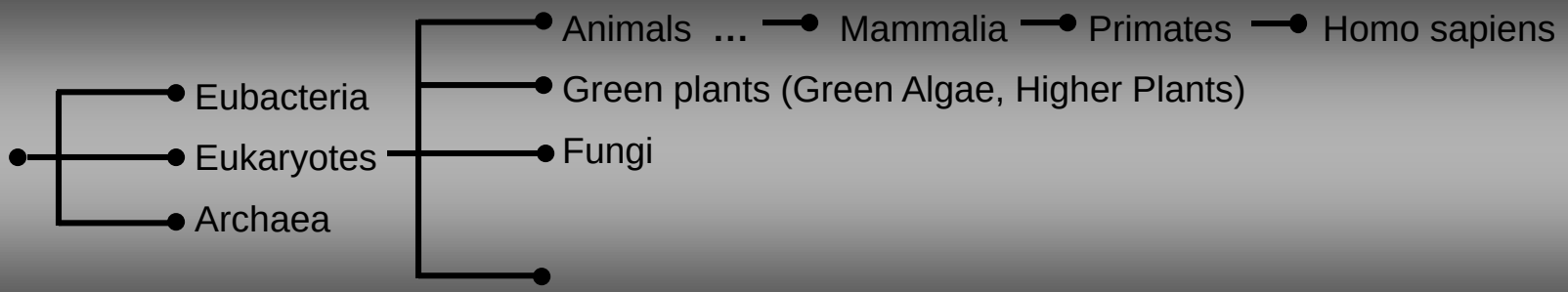
“Structure Computing”
How Things Got Started …



Tree of life web project :
http://tolweb.org/tree/phylogeny.html

“Structure Computing”
… and Where We Are Heading to





Why Conceptualize?

- Nomenclatures, thesauri, ontologies, …
- "Mapping problem" due to term variation
  - Natural language ⇨ domain knowledge
- "Structure computing" on knowledge structures
  - Lexical look-up
  - Relational navigation (general-specific, is-a)
  - Formal reasoning (inferencing)
- **Bio view: data annotation & data integration**

Bio View: Swiss-Prot and GO Terms

Entry information	
Entry name	**IL2_HUMAN**
Primary accession number	**P60568**
Secondary accession number	P01585
Integrated into Swiss-Prot on	July 21, 1986
Sequence was last modified on	July 21, 1986 (Sequence version 1)
Annotations were last modified on	March 21, 2006 (Entry version 29)

Name and origin of the protein	
Protein name	**Interleukin-2 [**
Synonyms	**IL-2**
Gene na	**e: IL2**
From	o sapiens

GO terms

GO:0005615; Cellular component: extracellular space *(traceable author statement)*.
GO:0005125; Molecular function: cytokine activity *(inferred from direct assay)*.
GO:0005134; Molecular function: interleukin-2 receptor binding *(traceable author sta*
GO:0019209; Molecular function: kinase activator activity *(traceable author stateme*
GO:0006916; Biological process: anti-apoptosis *(traceable author statement)*.
GO:0019735; Biological process: antimicrobial humoral response (sensu Vertebrata
GO:0007155; Biological process: cell adhesion *(traceable author statement)*.
GO:0007267; Biological process: cell-cell signaling *(traceable author statement)*.
GO:0006955; Biological process: immune response *(traceable author statement)*.
GO:0030101; Biological process: natural killer cell activation *(traceable author state*
GO:0042104; Biological process: positive regulation of activated T cell proliferation
GO:0030307; Biological process: positive regulation of cell growth *(traceable autho*
GO:0008284; Biological process: positive regulation of cell proliferation *(traceable a*
GO:0030217; Biological process: T cell differentiation *(traceable author statement)*.
QuickGo view.

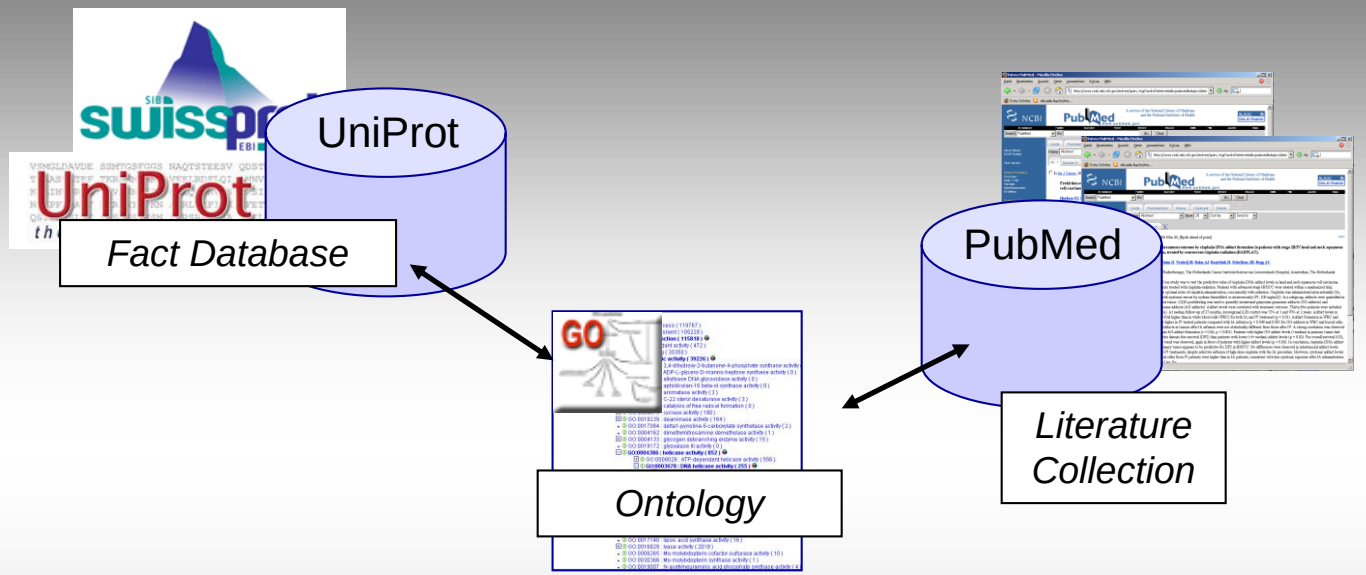
PubMed

Function:
required for T-cell proliferation and other activities crucial to the regulation of the immune response

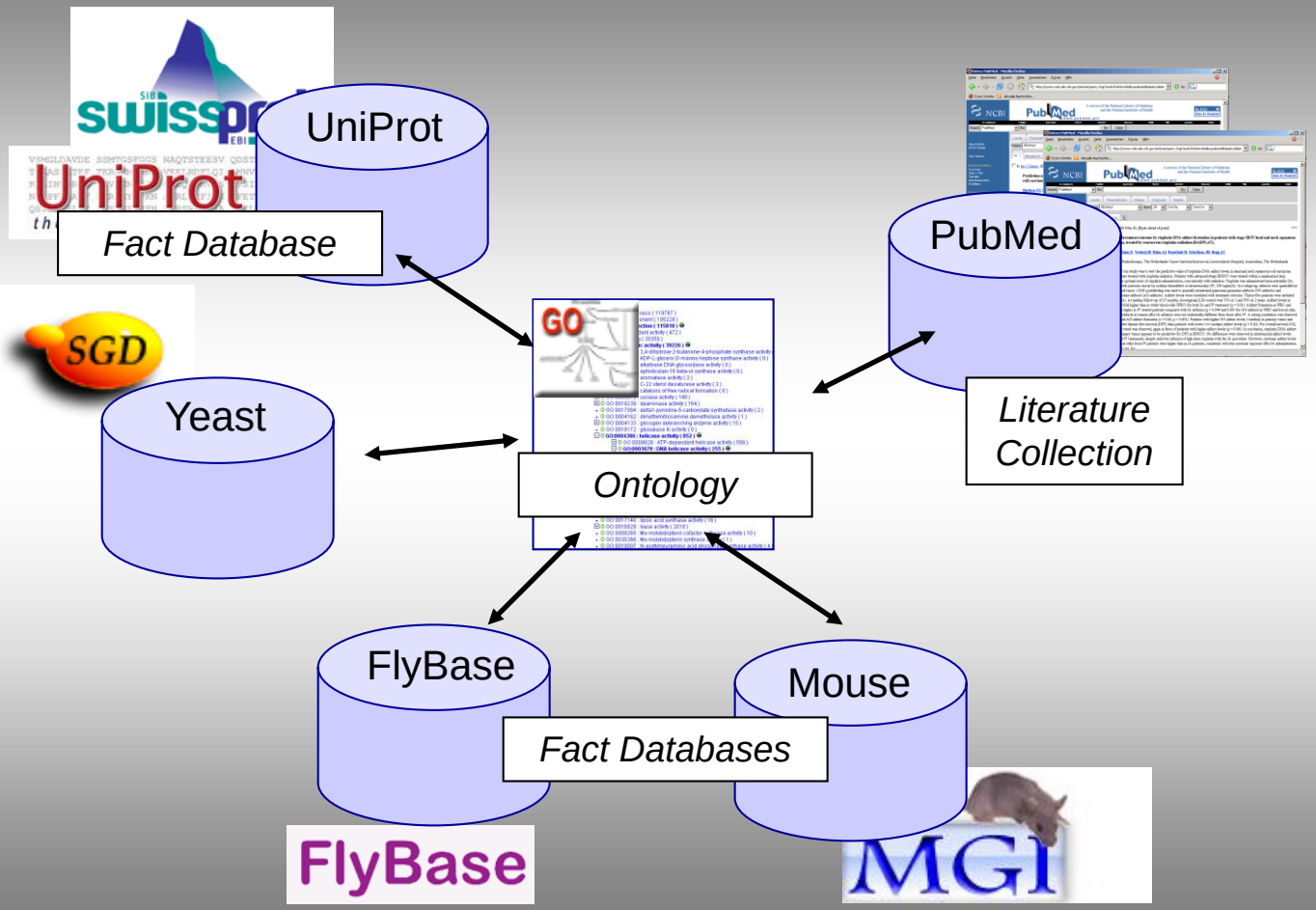
Location:
secreted protein

http://www.expasy.org/sprot/

Ontologies and Data Annotation



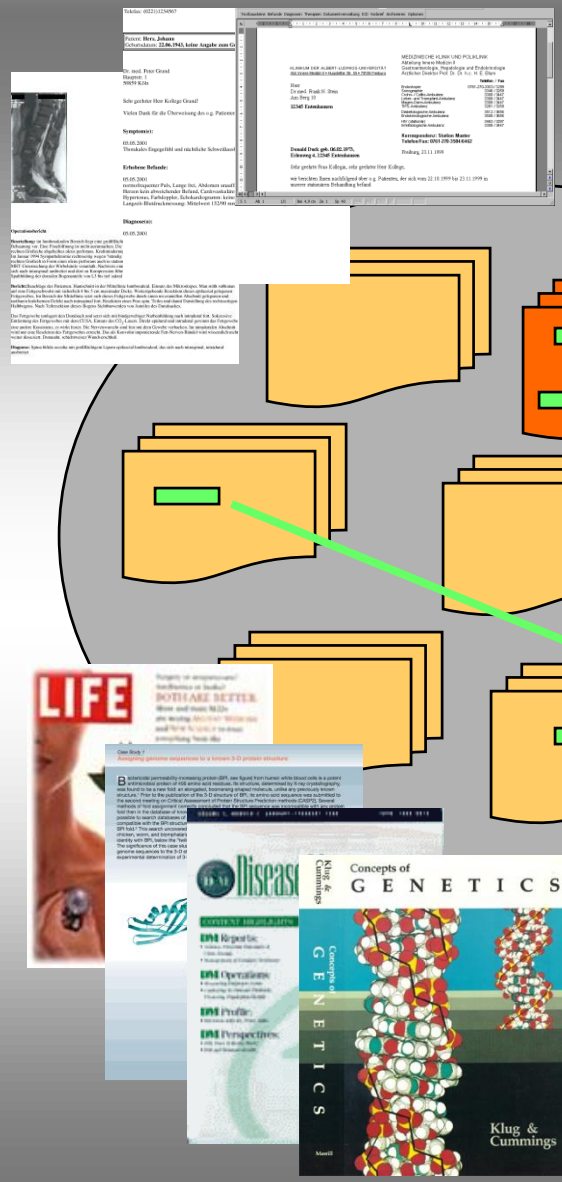
Ontologies and Data Integration



Why Conceptualize?

- Nomenclatures, thesauri, ontologies, …
- "Mapping problem" due to term variation
  - Natural language ⇨ domain knowledge
- "Structure computing" on knowledge structures
  - Lexical look-up
  - Relational navigation (general-specific, is-a)
  - Formal reasoning (inferencing)
- Bio view: data annotation & data integration
- **NLP view: text-based content management**
  - **Category classification (IR)**
  - **Semantic interpretation (IE, TM)**

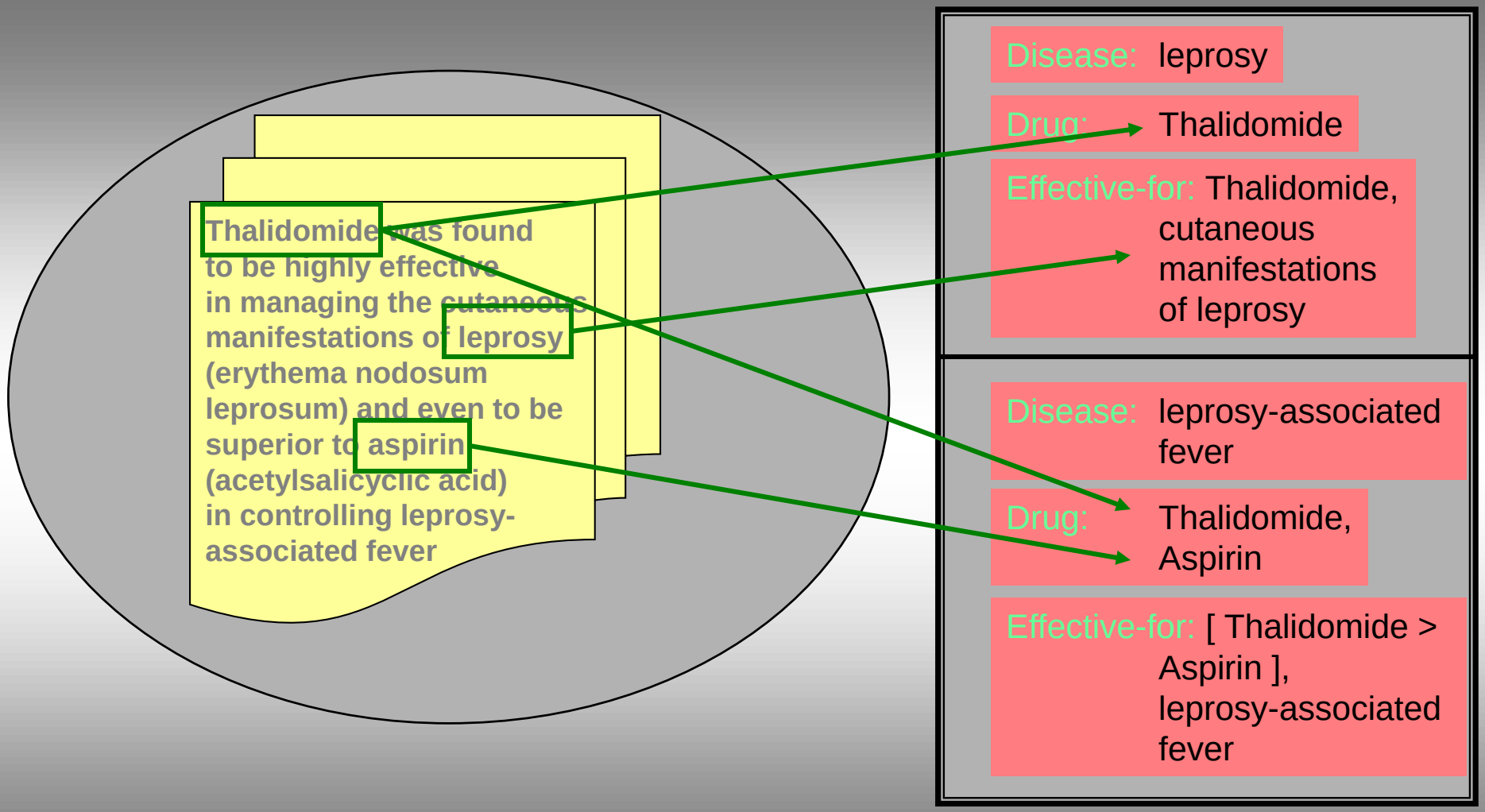
NLP view:
Two Text-Based CM Paradigms



Information Retrieval, Document Classification

Information Extraction, Text Mining

Information Extraction

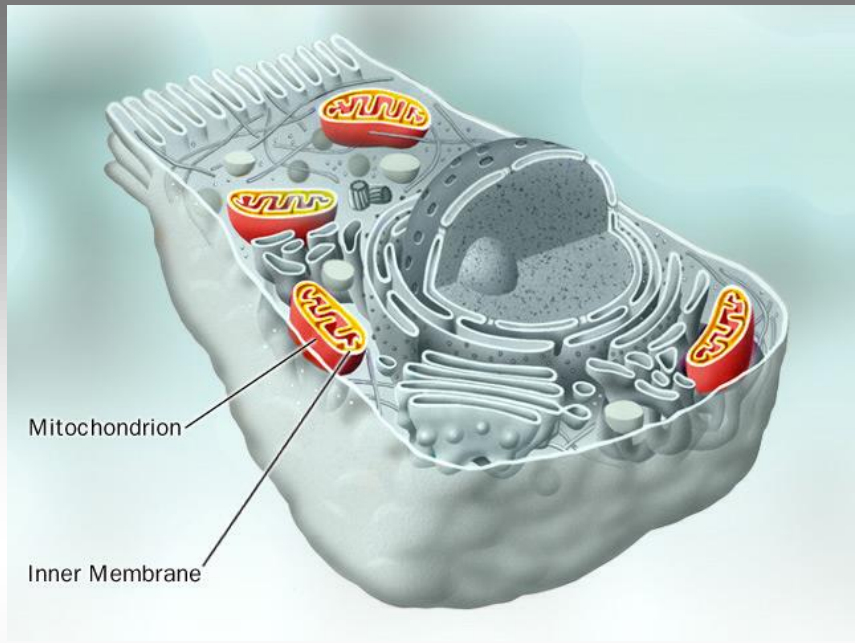


Ontologies for Information Extraction

S1 A mitochondrion provides the cell with energy in the form of ATP.

S2 The organelle possesses its own genetic material which is inherited maternally.

S3 The ATP synthesizing enzyme ATP synthase is located in the inner membrane.

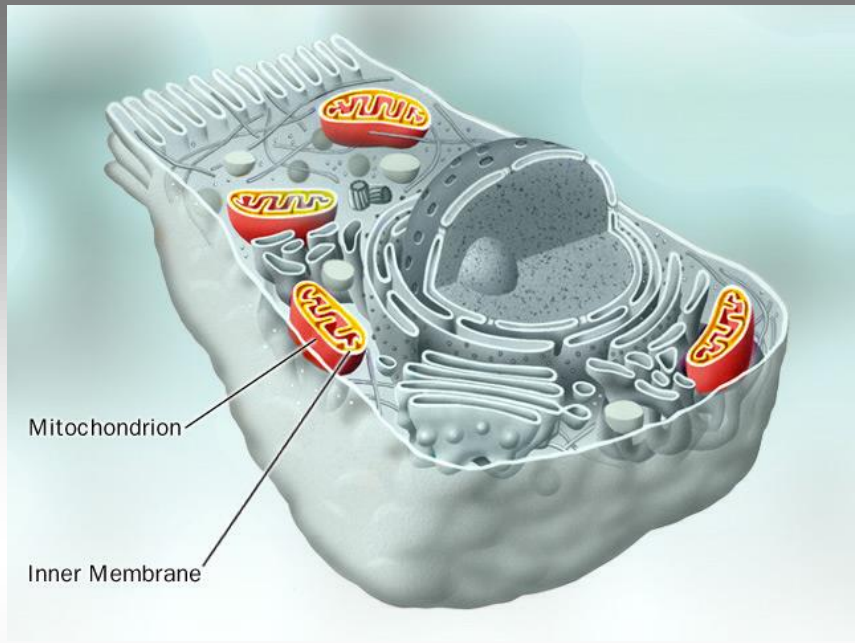


Ontologies for Information Extraction

S1 A mitochondrion provides the cell with energy in the form of ATP.

is-a

S2 The organelle possesses its own genetic material which is inherited maternally.



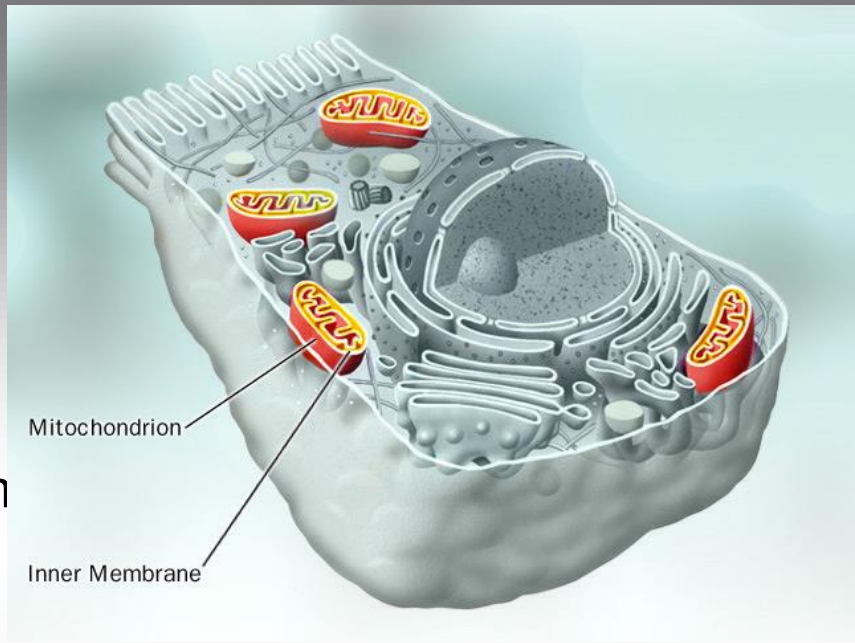
S3 The ATP synthesizing enzyme ATP synthase is located in the inner membrane.

Ontologies for Information Extraction

S1 A mitochondrion provides the cell with energy in the form of ATP.

S2 The mitochondrion possesses its own genetic material which is inherited maternally.

S3 The ATP synthesizing enzyme ATP synthase is located in the inner membrane.



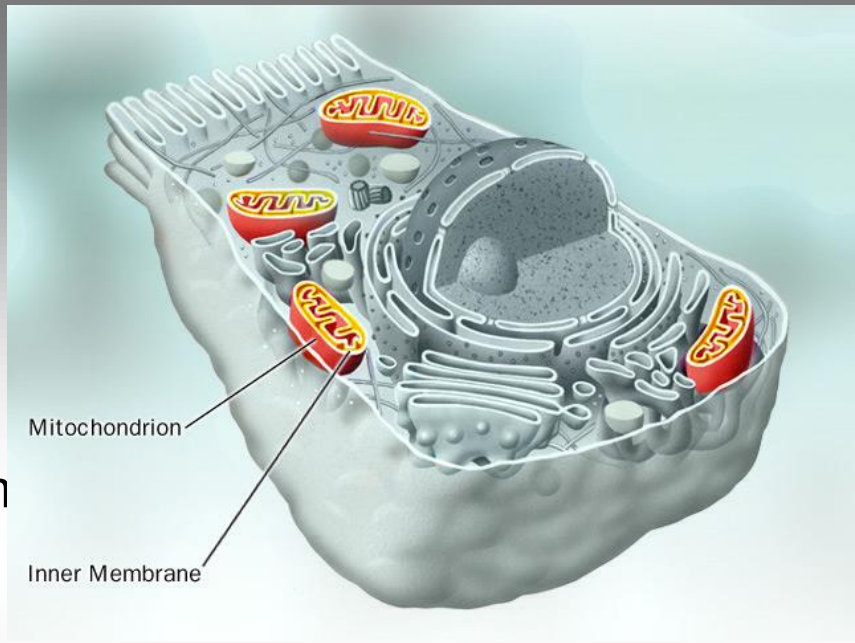
Ontologies for Information Extraction

S1 A mitochondrion provides the cell with energy in the form of ATP.

S2 The mitochondrion possesses its own genetic material which is inherited maternally.

part-of

S3 The ATP synthesizing enzyme ATP synthase is located in the inner membrane.



Conceptual Normalization

S1 A mitochondrion provides the cell with energy in the form of ATP.

S2 The mitochondrion possesses its own genetic material which is inherited maternally.

S3 The ATP synthesizing enzyme ATP synthase is located in the mitochondrial inner membrane.

Semantic Interpretation

"Normalized" Text Level

S1 A mitochondrion provides the cell with energy in the form of ATP.

S2 The mitochondrion possesses its own genetic material which is inherited maternally.

S3 The ATP synthesizing enzyme ATP synthase is located in the mitochondrial inner membrane.

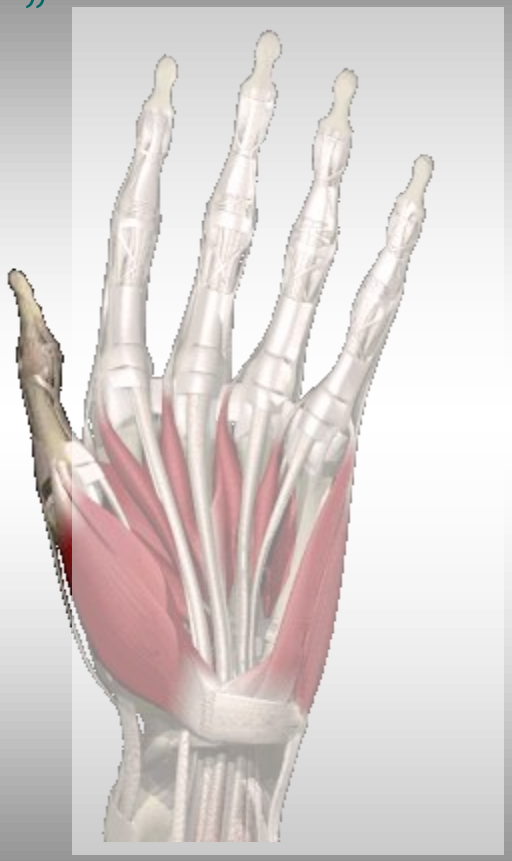
Propositional Level

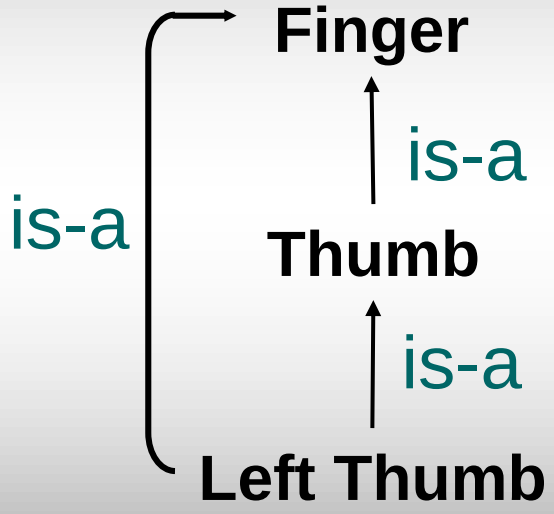
- **Provide [mitoch., cell, energy]**
- **Possess [mitoch., gen. material]**
- **Synthesize [ATP synthase, ATP]**
- **Located-in [ATP synthase, mitoch. inner membrane]**

Reasoning on Medical Ontologies

1. Taxonomy

„is-a“

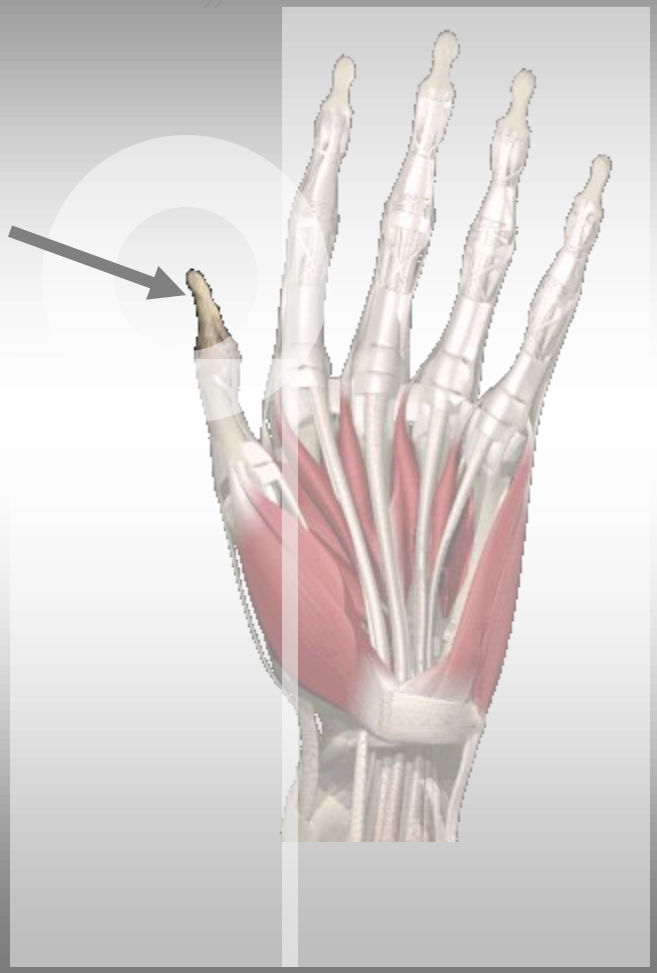




Reasoning on Medical Ontologies

2. Mereology

„part-of“



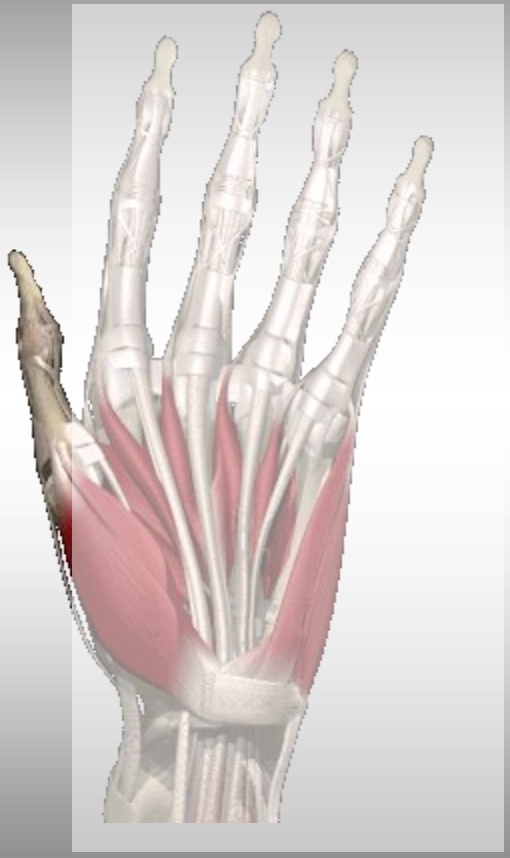
part-of

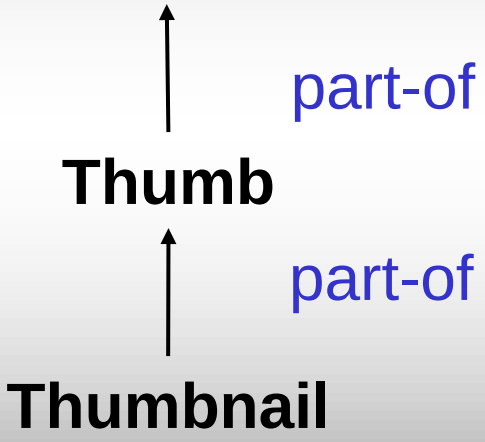
Thumbnail

Reasoning on Medical Ontologies

2. Mereology

„part-of“

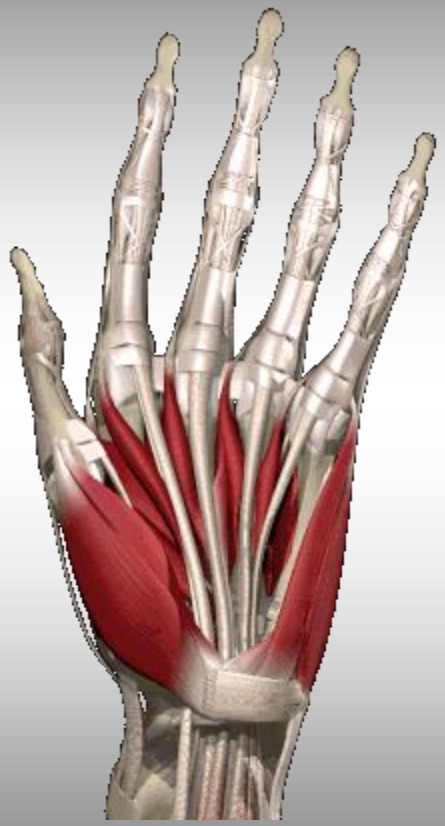


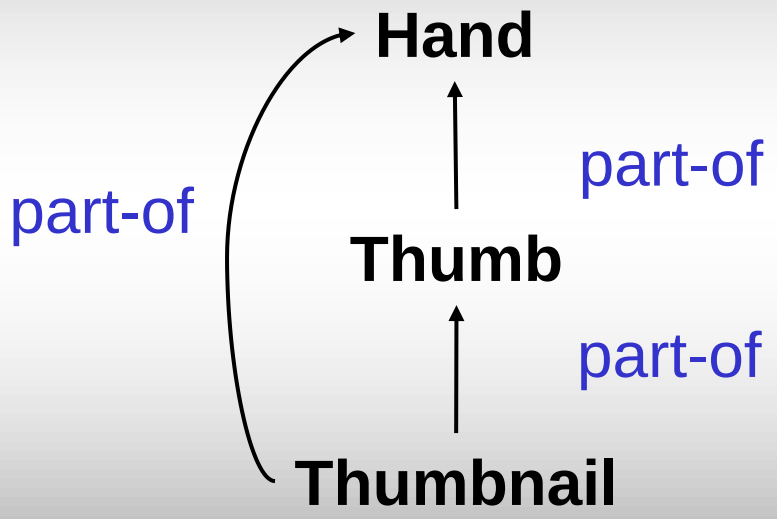


Reasoning on Medical Ontologies

2. Mereology

„part-of“

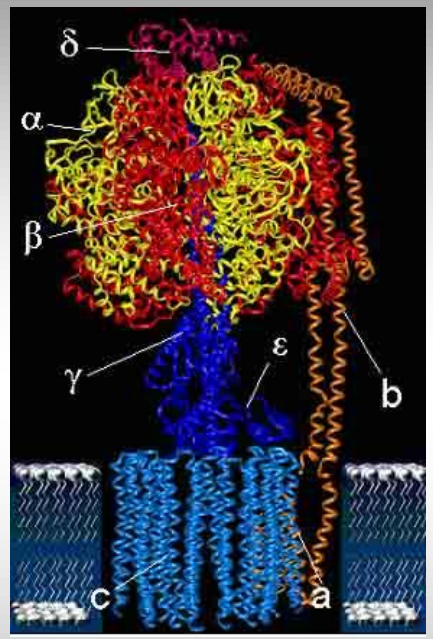


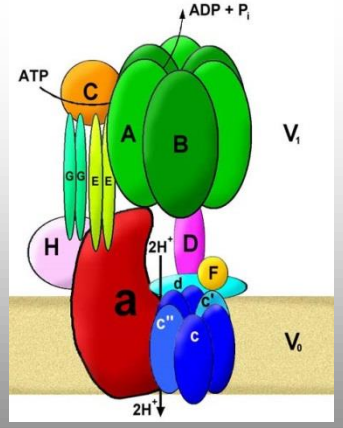


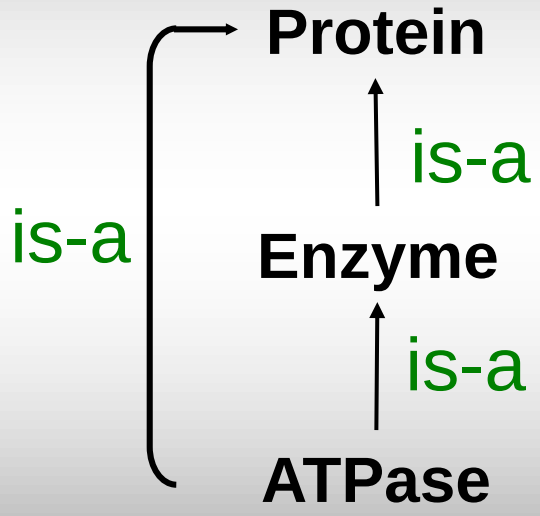
Reasoning on Bio Ontologies

1. Taxonomy

„is-a“



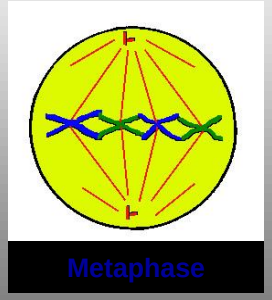




Reasoning on Bio Ontologies

1. Taxonomy
„is-a“

2. Mereology
„part-of“

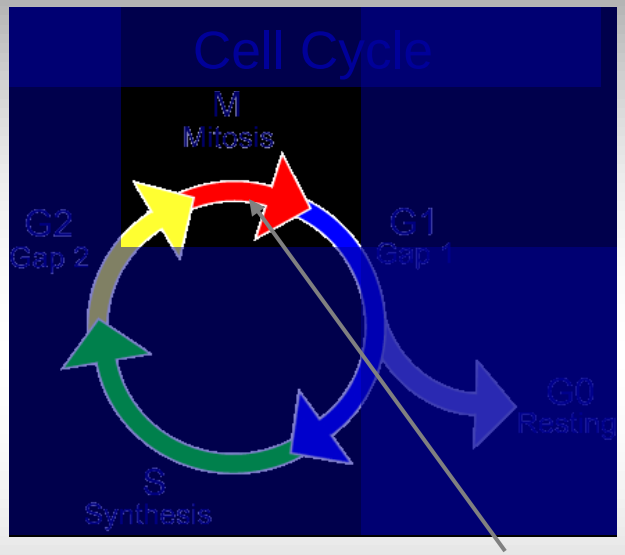


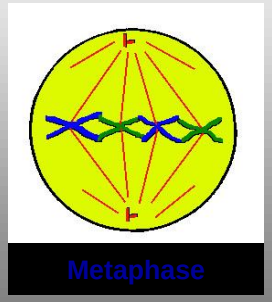
Metaphase

Reasoning on Bio Ontologies

1. Taxonomy
„is-a“

2. Mereology
„part-of“





Mitosis

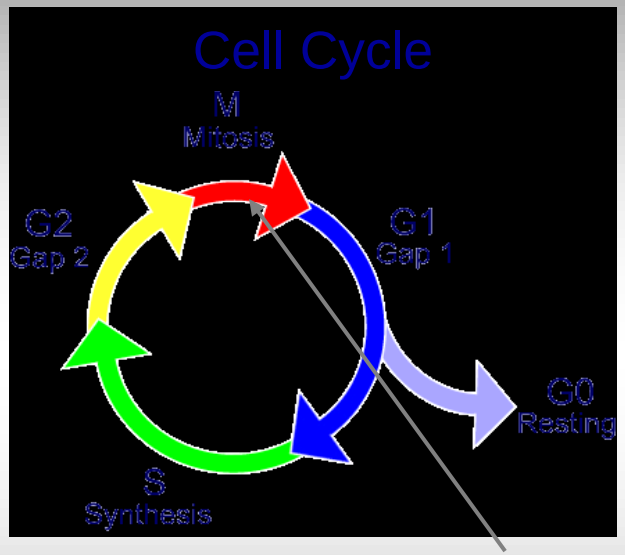
part-of

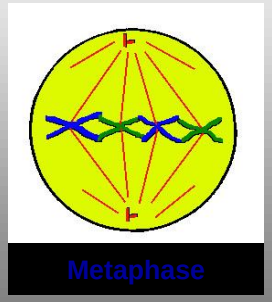
Metaphase

Reasoning on Bio Ontologies

1. Taxonomy
„is-a“

2. Mereology
„part-of“





Cell Cycle

part-of

Mitosis

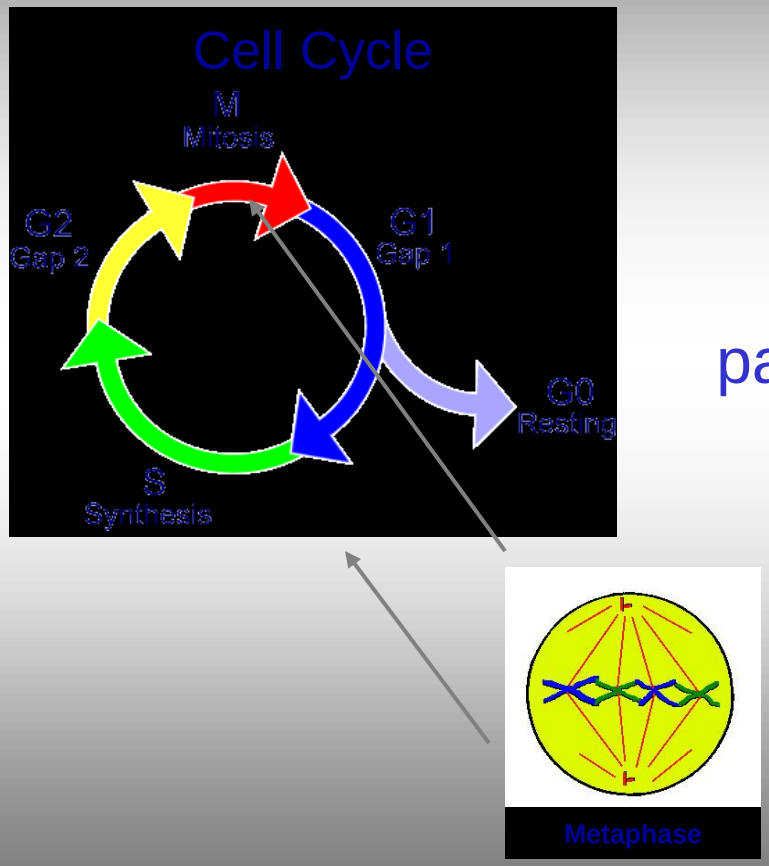
part-of

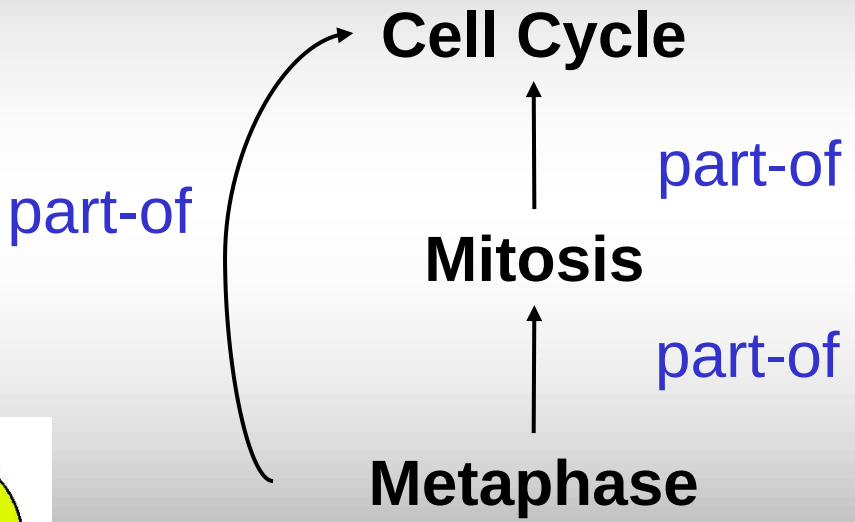
Metaphase

Reasoning on Bio Ontologies

1. Taxonomy
„is-a“

2. Mereology
„part-of“





… aber

- Die Hand ist Teil des Cellisten
- Der Cellist ist Teil des Orchesters

  ⇨ * Die Hand ist Teil des Orchesters

Lösung:

1. Es gibt unterschiedliche Subrelationen zu Teil-von (s. GermaNet!) und
2. Diese dürfen beim Folgern nicht gemischt werden

Relations Ontology (RO)

	name	transitive	symmetric	reflexive	anti-symmetric
foundational	is_a	+		+	+
	part_of	+		+	+
	integral_part_of	+		+	+
	proper_part_of	+			
	improper_part_of	+		+	
spatial	located_in	+		+	
	contained_in				
	adjacent_to				
temporal	transformation_of	+			
	derives_from	+			
	preceded_by	+			
participation	has_participant				
	has_agent				
	instance_of				

C `part-of` C_1 &
C_1 `has-part` C

Class Relations!

OWL – Web Ontology Language
http://www.w3.org/2001/sw/wiki/OWL

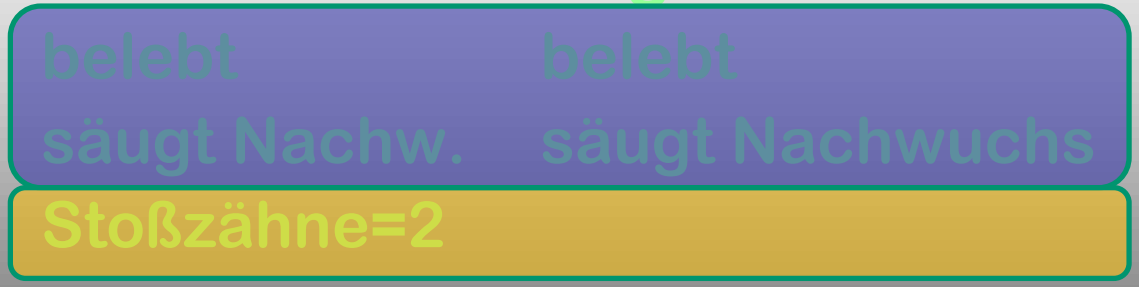
- Spezifikationsstandard für die Beschreibung von formalen Ontologien
  - Taxonomisches „Rechnen“ ⇨ Subsumption
- Formale Sprachbasis: Beschreibungslogik
  - Entscheidbare Teilmenge der Prädikatenlogik

Illustration der Subsumption

- Term_1 Term_2

⇨ Term_1 Is-a Term_2

- Elefant Säugetier



⇨ Elefant Is-a Säugetier

Technischer Kontext formaler Ontologien

http://www.w3.org/TR/2012/REC-owl2-primer-20121211/#OWL_2_EL

- Ontologie-Sprachen
  - formale Sprachen zur Repräsentation von Wissen
  - OWL-Dialekte
    - OWL2–EL (Beschreibungslogik-Variante = OWL 2 DL)
    - OWL2-QL (Datenbank-orientierte Anfragevariante)
    - OWL2-RL (effizienz-orientierte regelbasierte Variante für große RDF-Systeme)
- Ontologie-Inferenzmaschinen
  - automatisches Schließen (Inferenzen) über formal spezifiziertem Wissen
  - RACER, PELLET, FACT++, HERMiT …
- Ontologie-Entwicklungsumgebungen
  - PROTÉGÉ, SWOOP , NEON, …
- Ontologie-APIs
  - OWL-API, OWLink, Thea

DL (Description Logic) – OWL: Klassen-/Konzeptkonstruktoren

OWL Constructor	DL Syntax	Example
intersectionOf	$C_1 \sqcap \ldots \sqcap C_n$	Human $\sqcap$ Male
unionOf	$C_1 \sqcup \ldots \sqcup C_n$	Doctor $\sqcup$ Lawyer
complementOf	$\neg C$	$\neg$Male
oneOf	$\{x_1\} \sqcup \ldots \sqcup \{x_n\}$	{john} $\sqcup$ {mary}
allValuesFrom	$\forall P.C$	$\forall$hasChild.Doctor
someValuesFrom	$\exists P.C$	$\exists$hasChild.Lawyer
maxCardinality	$\leqslant nP$	$\leqslant$1hasChild
minCardinality	$\geqslant nP$	$\geqslant$2hasChild

http://www.cs.ox.ac.uk/ian.horrocks/Seminars/download/Horrocks_Ian_pt1.pdf

DL (Description Logic) – OWL: Ontologie-Axiome

OWL-Ontologie: Mischung aus Tbox- und Abox-Axiomen

OWL Syntax	DL Syntax	Example
subClassOf	$C_1 \sqsubseteq C_2$	Human $\sqsubseteq$ Animal $\sqcap$ Biped
equivalentClass	$C_1 \equiv C_2$	Man $\equiv$ Human $\sqcap$ Male
subPropertyOf	$P_1 \sqsubseteq P_2$	hasDaughter $\sqsubseteq$ hasChild
equivalentProperty	$P_1 \equiv P_2$	cost $\equiv$ price
transitiveProperty	$P^+ \sqsubseteq P$	ancestor$^+$ $\sqsubseteq$ ancestor

OWL Syntax	DL Syntax	Example
type	$a : C$	John : Happy-Father
property	$\langle a, b \rangle : R$	$\langle$John, Mary$\rangle$: has-child

http://www.cs.ox.ac.uk/ian.horrocks/Seminars/download/Horrocks_Ian_pt1.pdf

Fundamental Distinctions

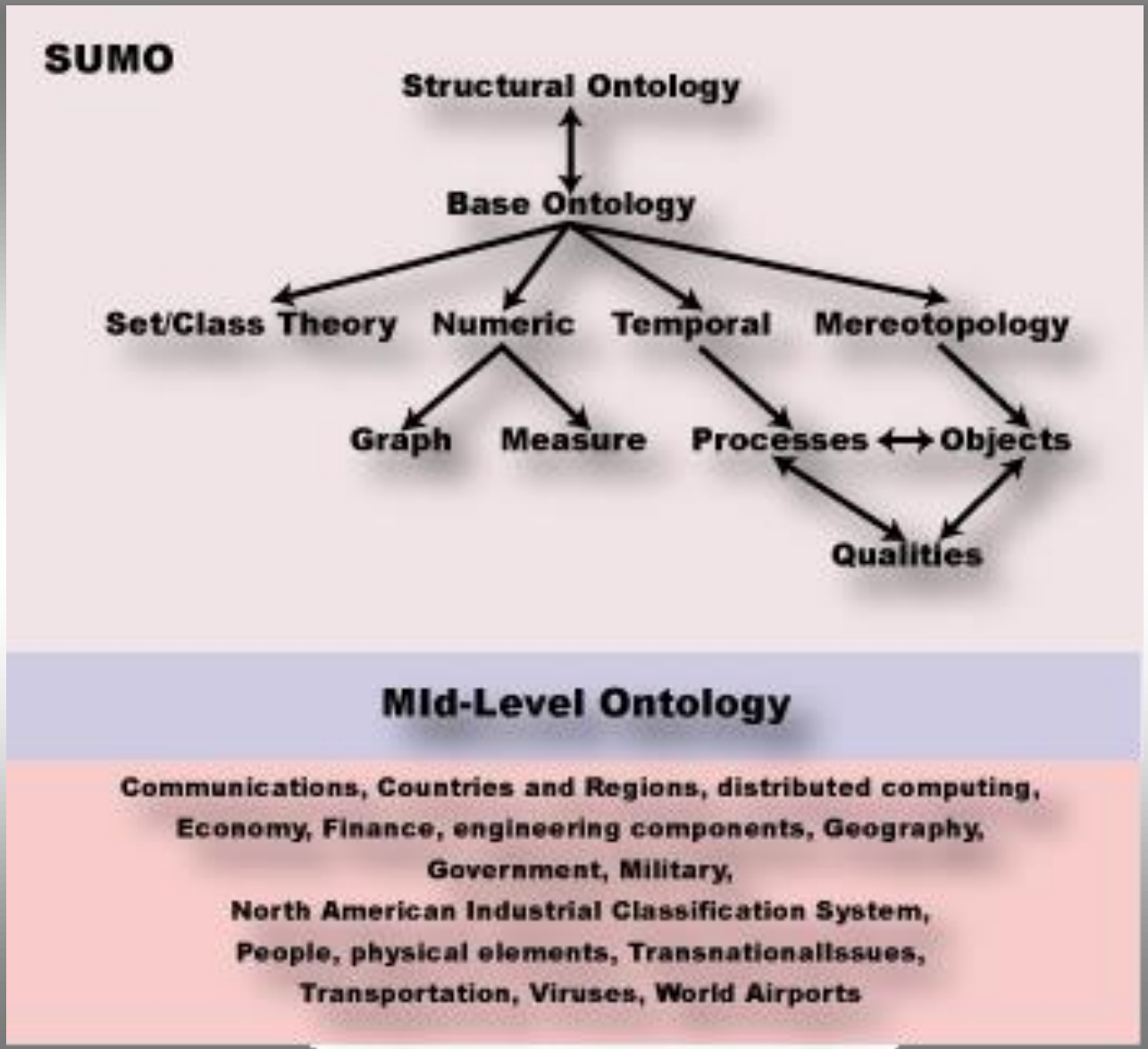
- Universals (classes, types, concepts) vs. particulars (instances, tokens, concrete & countable entities in the world which exist in space and time)
- Continuants (entities which endure, or continue to exist, through time while undergoing different sorts of changes)
  - e.g., molecule, cell, membrane, organ

  vs. occurrents (processes, events – entities which unfold themselves in successive temporal phases)
  - e.g., ion transport, cell division, breathing

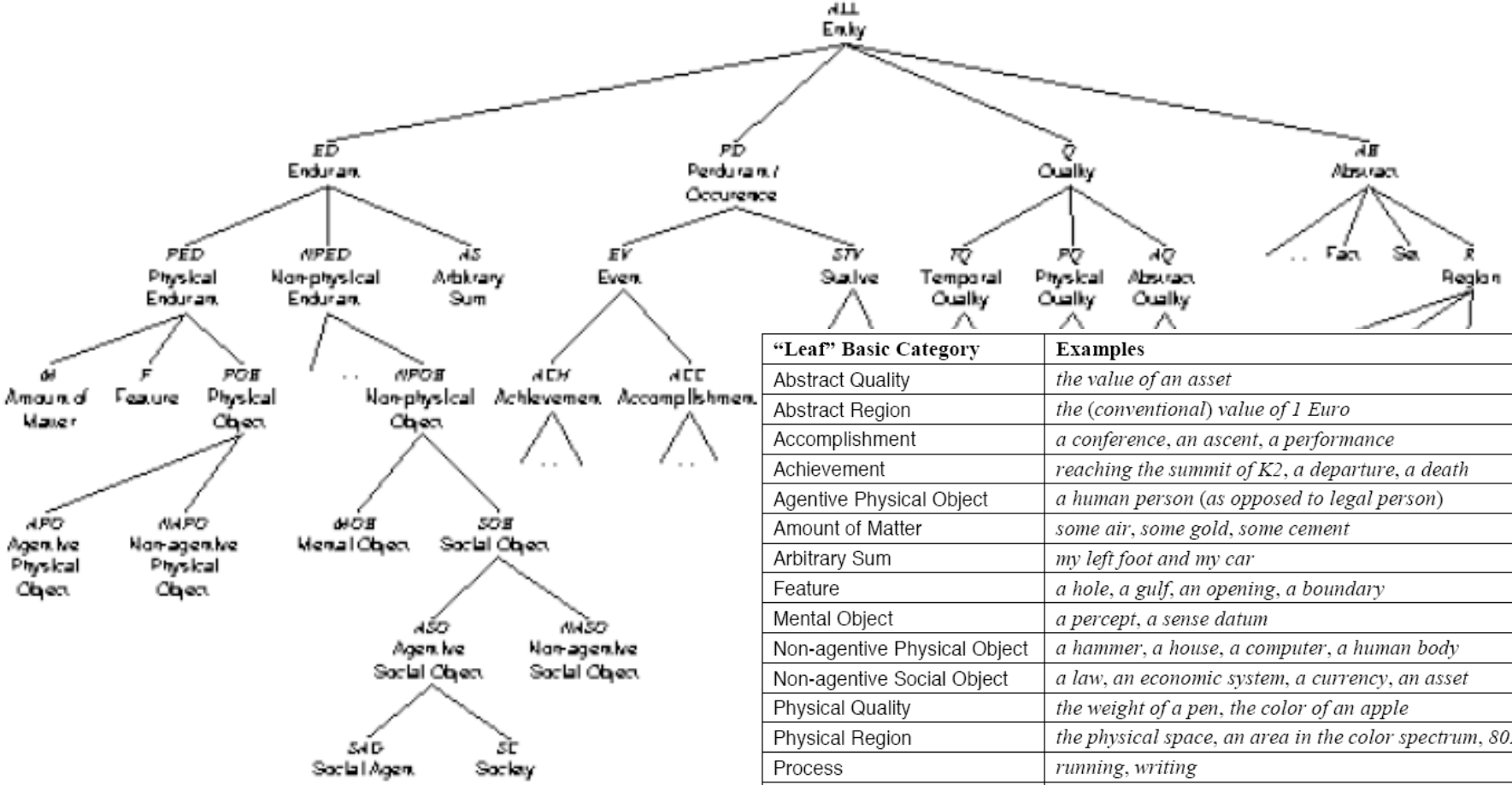
General Domain Ontologies

- SENSUS
  - 70.000 common-sense concepts
  - Extension and rearrangeent of WordNet
  - http://www.isi.edu/natural-language/resources/sensus.html
- OpenCyc
  - http://cyc.com/cyc/opencyc/overview
- SUMO (Suggested Upper Merged Ontology)
  - 20.000 common-sense concepts, 80,000 axioms
  - http://www.ontologyportal.org/
- DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering)
  - FOL
  - http://www.loa.istc.cnr.it/DOLCE.html

SUMO + Mid-Level Ontology Layer



Upper Ontologies
DOLCE



"Leaf" Basic Category	Examples
Abstract Quality	*the value of an asset*
Abstract Region	*the (conventional) value of 1 Euro*
Accomplishment	*a conference, an ascent, a performance*
Achievement	*reaching the summit of K2, a departure, a death*
Agentive Physical Object	*a human person (as opposed to legal person)*
Amount of Matter	*some air, some gold, some cement*
Arbitrary Sum	*my left foot and my car*
Feature	*a hole, a gulf, an opening, a boundary*
Mental Object	*a percept, a sense datum*
Non-agentive Physical Object	*a hammer, a house, a computer, a human body*
Non-agentive Social Object	*a law, an economic system, a currency, an asset*
Physical Quality	*the weight of a pen, the color of an apple*
Physical Region	*the physical space, an area in the color spectrum, 80Kg*
Process	*running, writing*
Social Agent	*a (legal) person, a contractant*
Society	*Fiat, Apple, the Bank of Italy*
State	*being sitting, being open, being happy, being red*
Temporal Quality	*the duration of World War I, the starting time of the 2000 Olympics*
Temporal Region	*the time axis, 22 june 2002, one second*

Table 1. Examples of "leaf" basic categories.

Conclusions

- No ontology without application!
- Formal foundations for domain modeling and ontology engineering
- Ontology configuration
  - Upper level ▽ domain ontology
  - Domain ontology: composition of departmental (subdomain) ontologies
  - Departmental ontology: hierarchy of different grain sizes of each department
- Domain Knowledge Base: Ontology linked to a Fact Data Base (assertions)

Recommended Readings

- **Computational Lexicons**
  - C. Fellbaum (Ed.) (1998), *WordNet: An Electronic Lexical Database*. MIT Press.
  - J. Ruppenhofer, M. Ellsworth, M.R. Petruck, C.R. Johnson & J. Scheffcyzk (2006), *FrameNet II: Extended Theory and Practice*. (e-book @ FrameNet website)
- **Ontologies**
  - S. Staab & R. Studer (Eds.) (2004), *Handbook on Ontologies*. Springer Publisher