13 August 2024 09:19

-> Hyponymy :

-> car is a hyponym of vehicle.

-> dog is a hyponym of animal.

-> mango is a hyponym of fruit.

-> Hypernymy : Conversely superordinate.

-> vehicle is a hypernym of car.

-> Hyponymy more formally :

-> Entailment : Sense A is a hyponym of sense B if being an A entails being a B.

Eg : Dog & Animal.

-> Transivity : A hyponym B & B hyponym of C entails A hyponym C.

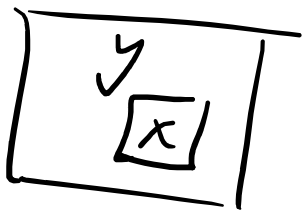
Eg. Clark Kent Superman. Kryptonian.

Eg: Clark Kent, Superman, Kryptonian.

-> Meronyms & Holonyms:

-> Meronymy: Asymmetric & Transitive

X is a meronym of Y if it denotes a part of Y.



-> The inverse relation is holonymy.

-> Meronym Holonym

-> Porch -> House

The porch is meronym of House &

House is holonym of Porch.

-> Wheel -> Car

-> Leg -> Chair

-> Nose -> Face

→ WordNet:

→ A hierarchically organized lexical DB.

→ A machine-readable thesaurus & aspects of a dictionary.

POS	No. of Synsets
Noun	82115
Verb	13767
Adj	18156
Adv	3621

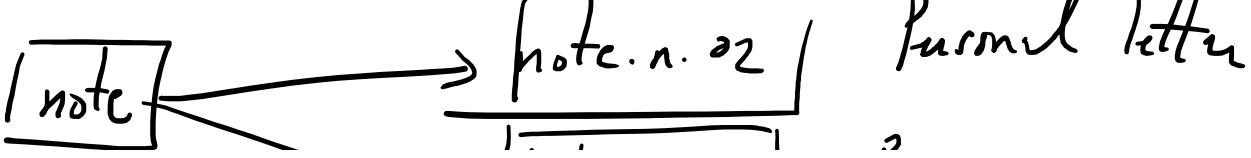
→ Synsets: Set of synonyms representing a sense.

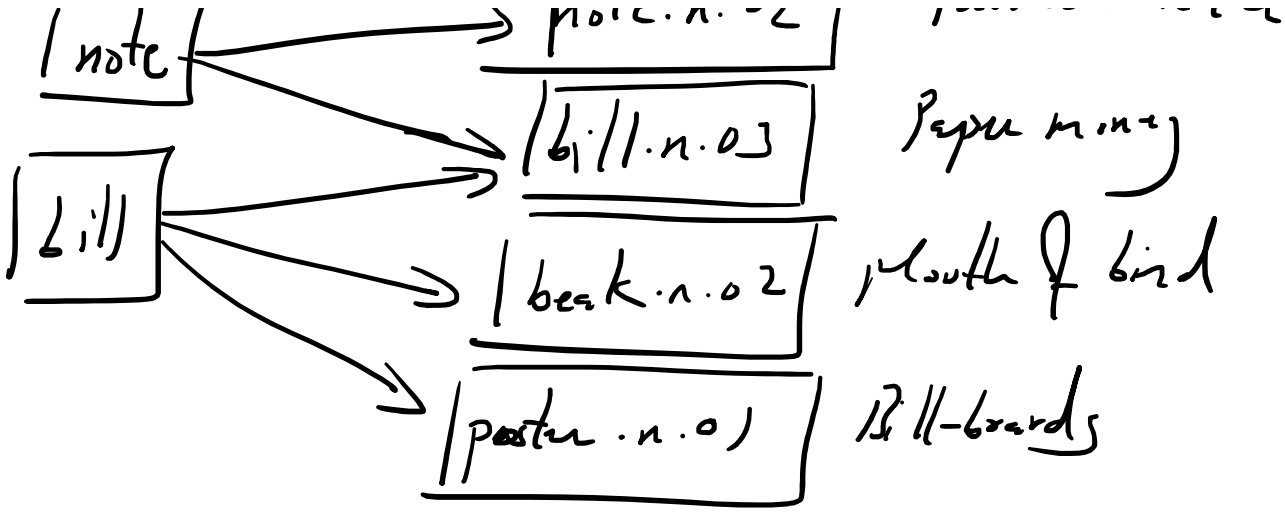
Eg: Chump as a noun: Person who is gullible.

{ chump[1], fool[2], gull[3], patsy[4], fall-guy[5], sucker[6] }

→ Each of these senses share same glossary.

→ Lemma vs Synsets:





→ Wordnet Hierarchies:

→ Synonyms / Hypernyms:

S1:

mouse : rodent, gnawer

↳ placental mammal, eutherian mammal

↳ mammal

↳ vertebrate, craniate

↳ chordate

↳ animal

↳ organism

↳ living thing

↳ whole

↳ object

↳ physical entity
↳ entity.

S2 : mouse, computer mouse.
↳ electronic device
↳ device
↳ instrument
↳ artifact
↳ units
↳ objects
↳ entity.

→ Word Similarity :

-> Synonymy is a binary relation.
-> Looser metric (Word similarity or word distance).
-> Relations b/w senses :
↳ Bank[1] is similar to fund[3]
↳ Bank[2] is similar to slope[5]

Bank —(x)→ Fund

Bank ≠ Fund

S1 - Economics ⟶ S1 - Economics

S2 - River ⟶ ~~=~~

-> Distributional Algorithms:

-> Thesaurus-Based Algorithms:

-> Thesaurus Based Word Similarity:

-> Subsumption & Hypernymy Hierarchy.

-> Similar words are near-synonyms.

Eg: car, gasoline : related, but not similar.

car, bicycle : similar.

-> Path-Based Similarity:

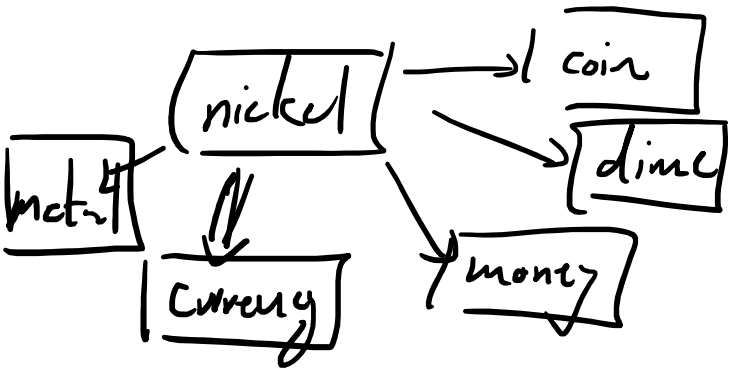
-> Two words are similar if they are nearby in the hypernymy graph.

-> pathlen(c1, c2) = number of edges in shortest path (in hypernymy graph) b/w senses c1 & c2.

-> $Sim_{path}(c_1, c_2) = \frac{1}{1 + pathlen(c_1, c_2)}$

-> $Sim(w_1, w_2) = \max_{c_1 \in senses(w_1),\ c_2 \in senses(w_2)} Sim(c_1, c_2)$

-> Shortest Path in the Hierarchy :



-> Leacock - Chodorow (L-C) Similarity

-> $Sim_{LC}(c_1, c_2) = -\log(pathlen(c_1, c_2) / 2d)$

d : Maximum Depth of Hierarchy Tree.

-> Problems :

Lengths to different edges.

-> Concept Probabilities Model :

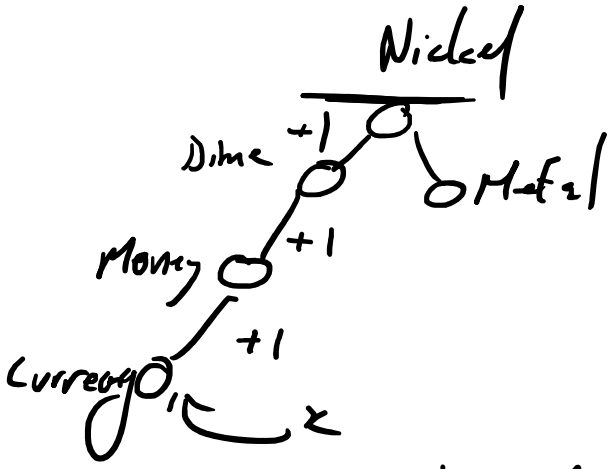
-> For each concept (synsets) c, let P(c) be the prob that a randomly selected word in a corpus is an instance (hyponym) of c.

→ $P(Root) = 1$

→ Estimating:

→ Counting "concept activations" in a corpus -

→ Each occurrence counts.



→ Information Content: Axiomatic

$$IC(c) = -\log P(c)$$

↑ concept ≈ sense of word

→ Lowest Common Sub-Summer

$LCS(c_1, c_2)$ = the lowest node in the hierarchy tree that subsumes (hypernym) of c_1 & c_2.

→ Information Content (IC) as a similarity metric.

-> Information Content "IC" as a similarity metric.

-> Resnik Similarity: How similar 2 words are depends on how much they have in common.

-> $Sim_{resnik}(C_1, C_2) = IC(LCS(C_1, C_2))$

$= -\log p(LCS(C_1, C_2))$

-> Lin - Similarity:

Lin:

$$Sim_{Lin}(C_1, C_2) = \frac{2 \log p(LCS(C_1, C_2))}{\log p(C_1) + \log p(C_2)}$$

-> Jiang - Conrath Distance (JC Similarity)

↳ Assigns lengths to graph edges.

$$-> Sim_{JC}(C_1, C_2) = \frac{1}{JC(C_1) + IC(C_2) - 2 \times IC(LCS(C_1, C_2))}$$