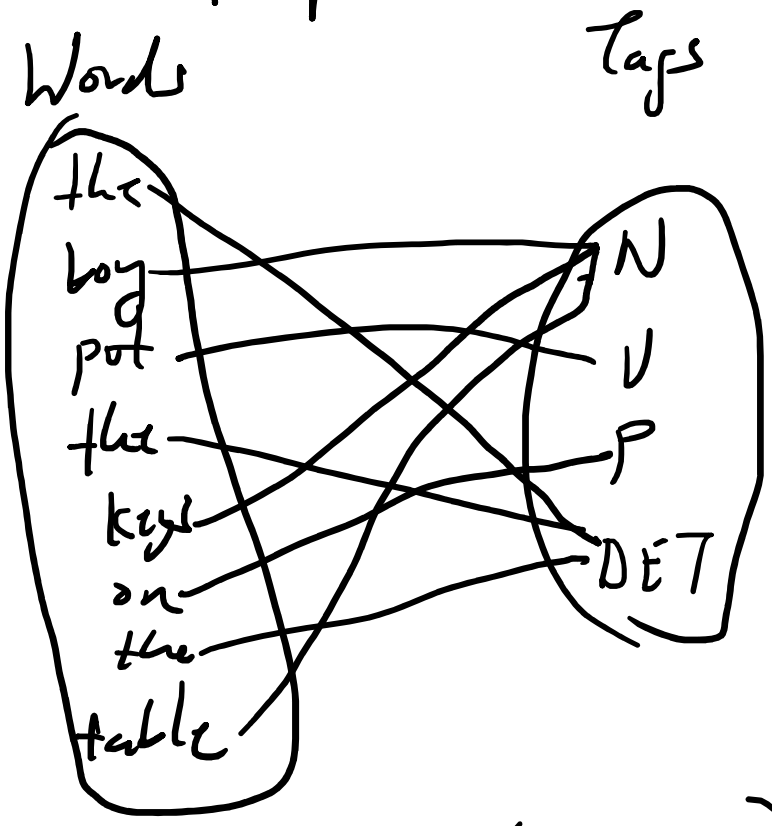
30 July 2024 09:21

→ Disambiguity : Race ~ Semantic Analysis.

→ Parts-of-Speech (POS)



→ Open Class Words (Content)
  ↳ Nouns, Verbs, Adjectives, Adverbs.

→ Closed Class Words
  ↳ Pronouns, Determiners, Prepositions, Connectives

→ N - Noun PRO - Pronoun
V - Verb

-> PRO - Pronoun

V - Verb

DET - Determiner.

ADJ - Adjective

ADV - Adverb

P - Preposition

-> The grand jury commented on a number of other topics.

-> The/DT -> grand/JJ -> jury/NN -> commented/VBD -> on/IN -> a/DT -> number/NN -> of/IN -> other/JJ -> topics/NNS.

-> Back

-> The back door. ~ JJ

-> On my back. -> NN

-> Win the votes back. -> RB

-> Promised to back the bill -> VB

-> the word itself :

-> arrows

-> Like flies

-> Local Context : (the a) the bear

run/drink the grizzly bear

-> Rule Based Approach :

-> Statistical Tagging : (TBL Tagger)

-> Probabilistic : T/ W

-> TBL Tagger :

-> the can was rusted.

The/DT -> can/MD -> was/VBD -> rusted/VBD.

-> MD -> NN : DT/_

-> VBD -> VBN : VBD/_

-> Probabilistic Tagging :

Generative Model : (Joint)

Discriminative Mode : (Conditional)

-> P(d, c) in terms of P(d|c)

-> Naive Bayes, Hidden Markov Model ... etc

-> Maximum Entropy, CRF ... DM.

-> Probabilistic Tagging :

W = $w_1 \ldots w_n$ ~ Observed

T = $t_1 \ldots t_n$ ~ (Unknown)

$\hat{T} = \text{argmax}_T \; P(T|W)$

$= \text{argmax}_T \; P(t_1 \ldots t_n | w_1 \ldots w_n)$

$\text{argmax} \; P(W|T) \times P(T)$

↳ $\text{argmax}_T \quad \frac{P(W|T) \times P(T)}{P(W)}$

$$\text{argmax}_T \quad P(W|T) \times P(T)$$

$$= \text{argmax}_T \quad \prod_i^n P(w_i | w_1 \ldots w_{i-1}, t_1 \ldots t_i) \times P(t_i | t_1 \ldots t_{i-1})$$

→ Bigram Assumption:

$$P(t_i | t_1 \ldots t_{i-1}) \approx P(t_i | t_{i-1})$$

∴ Using Simplification

$$\hat{T} = \text{argmax}_T \prod_i^n P(w_i | t_i) \times P(t_i | t_{i-1})$$