

---

---

# NLP Lecture 3

3 hours

---

---

# Wordnet

- WordNet is the lexical database i.e. dictionary for the English language, specifically designed for natural language processing.
- Synset is a special kind of a simple interface that is present in NLTK to look up words in WordNet. Synset instances are the groupings of synonymous words that express the same concept.
- Example -

```
from nltk.corpus import wordnet
for syn in wordnet.synsets("good"):

    for l in syn.lemmas():

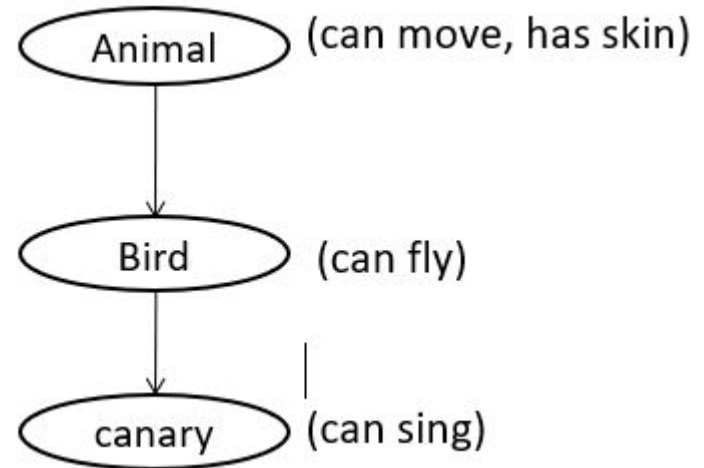
        synonyms.append(l.name())
```

```
{'beneficial', 'just', 'upright', 'thoroughly', 'in_force', 'well', 'skilful',
'skillful', 'sound', 'unspoiled', 'expert', 'proficient', 'in_effect',
'honorable', 'adept', 'secure', 'commodity', 'estimable', 'soundly',
'right', 'respectable', 'good', 'serious', 'ripe', 'salutary', 'dear',
'practiced', 'goodness', 'safe', 'effective', 'unspoilt', 'dependable',
'undecomposed', 'honest', 'full', 'near', 'trade_good'}
```

# Psycholinguistic Theory

- Human lexical memory for nouns as a hierarchy.
- *Can canary sing?* - *Pretty fast response.*
- *Can canary fly?* - *Slower response.*
- *Does canary have skin?* – *Slowest response*

Wordnet - a lexical reference system based on psycholinguistic theories of human lexical memory.



# Lexical matrix

Word Meanings	Word Forms				
	F1	F2	F3		Fn
M1	E1,1	E1,2			
M2		E2,2			
M3			E3,3		
Mm					Em,n

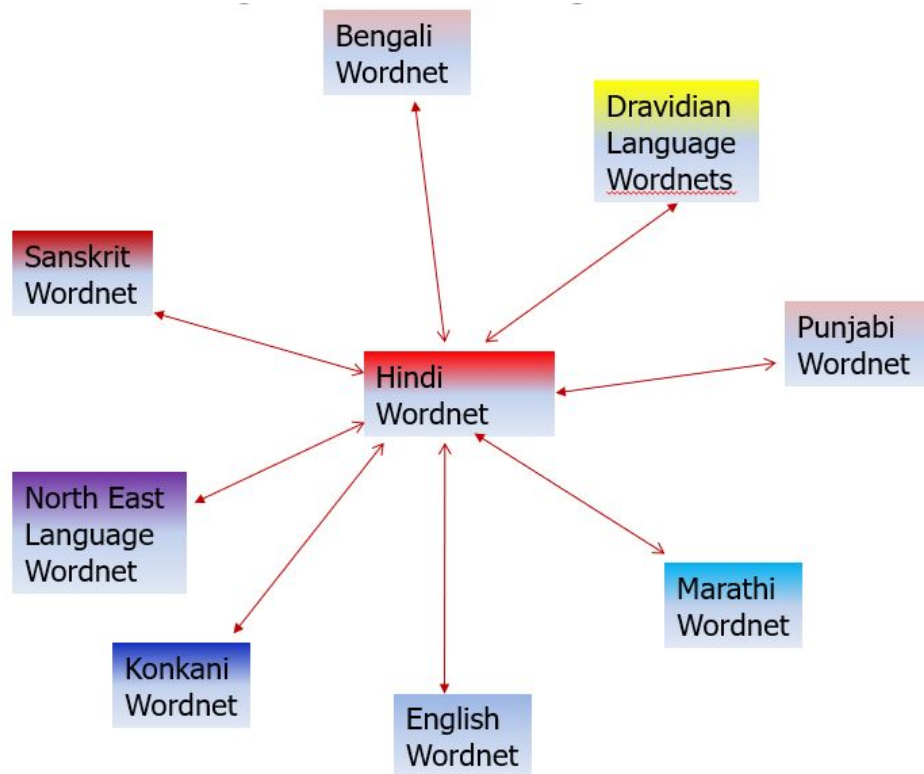
Synonymous Words

Polysemous Words

# Wordnet - Lexical Matrix (with examples)

Word Meanings	Word Forms				
	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	...	F <sub>n</sub>
M <sub>1</sub>	(depend) E <sub>1,1</sub>	(bank) E <sub>1,2</sub>	(rely) E <sub>1,3</sub>		
M <sub>2</sub>		(bank) E <sub>2,2</sub>		(embankment) E <sub>2,...</sub>	
M <sub>3</sub>		(bank) E <sub>3,2</sub>	E <sub>3,3</sub>		
...				...	
M <sub>m</sub>					E <sub>m,n</sub>

# Linked Wordnets in India



# Stated Fundamental Application of Wordnet: Sense Disambiguation

Determination of the correct sense of the word

*The crane ate the fish vs.*

*The crane was used to lift the load*

*bird vs. machine*

# The problem of Sense tagging

Given a corpora To Assign correct sense to the words.

This is sense tagging. Needs Word Sense Disambiguation (WSD)

Highly important for Question Answering, Machine Translation, Text Mining tasks.



# Word Sense Disambiguation (WSD)

For example, consider the two sentences.

“The bank will not be accepting cash on Saturdays. ”

“The river overflowed the bank.”

The word bank in the first sentence refers to the commercial (finance) banks, while in second sentence, it refers to the river bank.

<https://towardsdatascience.com/a-simple-word-sense-disambiguation-application-3ca645c56357>

# Query Expansion

- Query expansion (QE) is the process of reformulating a given query to improve retrieval performance in information retrieval operations, particularly in the context of query understanding.

<https://github.com/ellisa1419/Wordnet-Query-Expansion/blob/master/wordnet.py>

