Word2Vec → Tomas Mikolov → 2013 → Google

⇓ Community support

Google ↓ Facebook → Fast text ⇓ Supports multiple languages

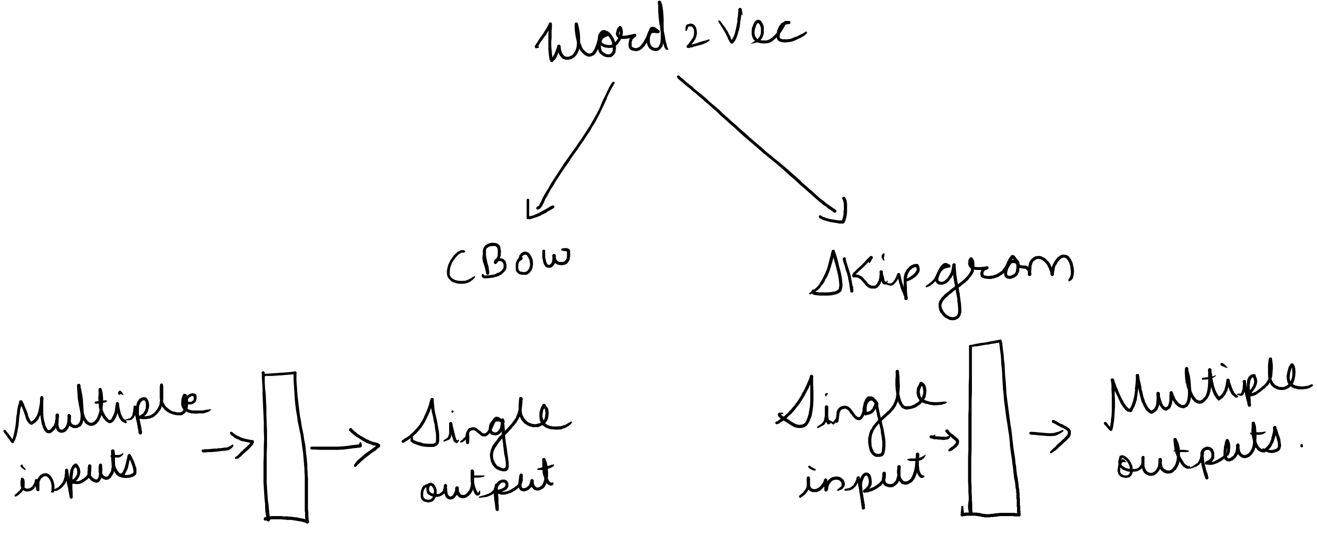
We can use Word2Vec in 2 ways:

① Pretrained model → Download → use

→ 3 Billions words → Google news dataset

② Customised model →

Architectures of Word2Vec :→



Word2Vec

CBOW

Skipgram

Multiple inputs → → Single output

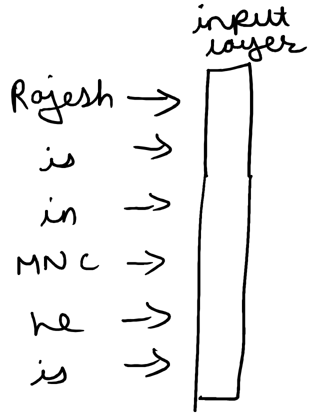
Single input → → Multiple outputs.

CBow → Continuous Bag of words

"Rajesh is working in MNC, he is a hardworking guy."

⇓ Target word

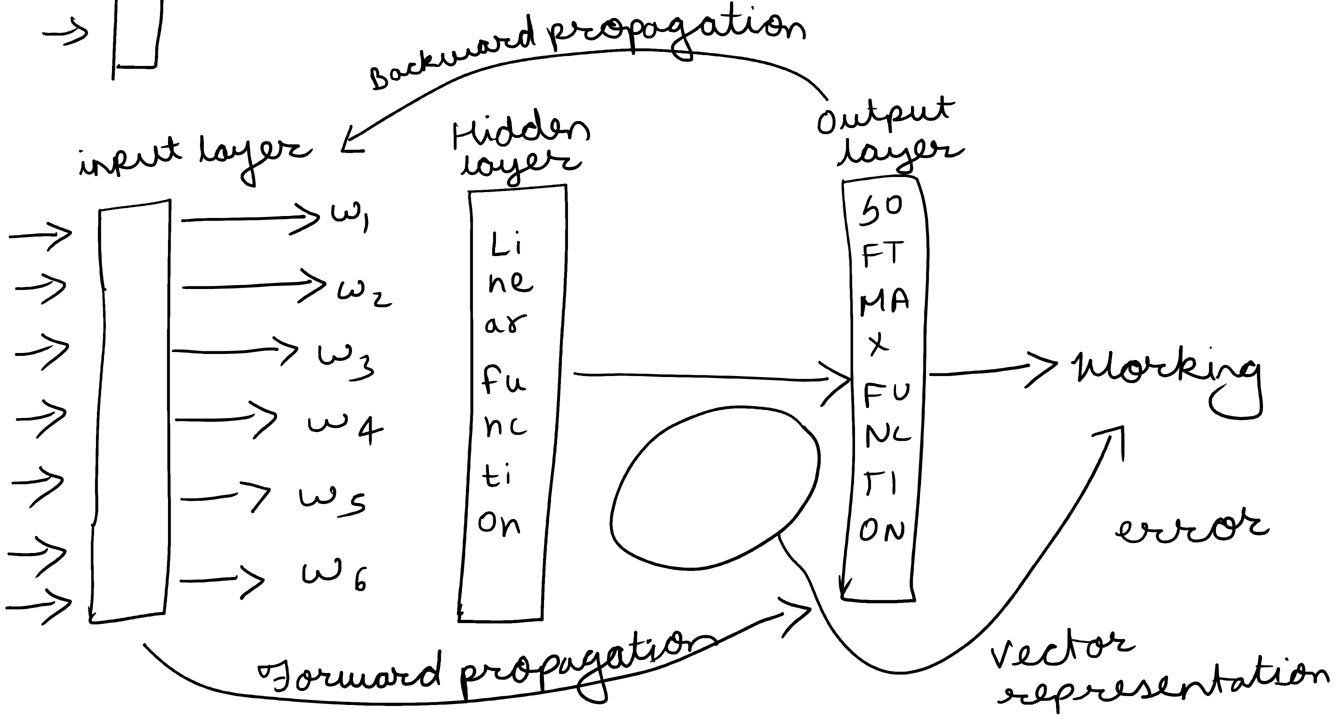
⇓ contextual words



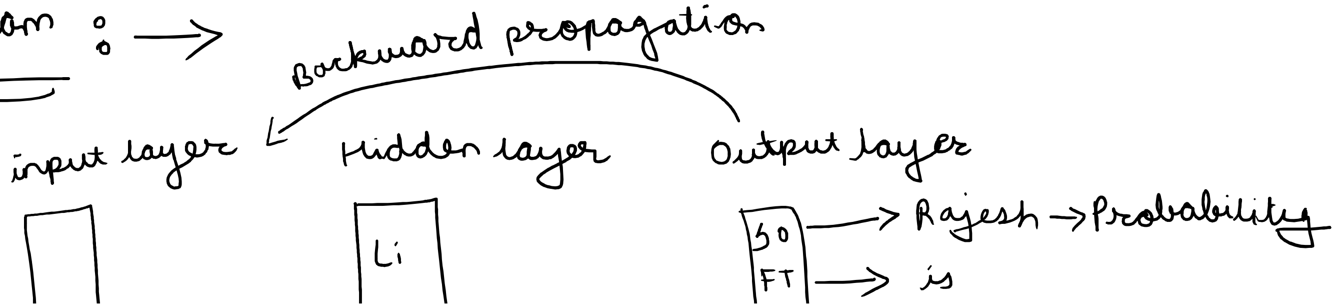
Rajesh = [1, 0, 0, 0, ----]

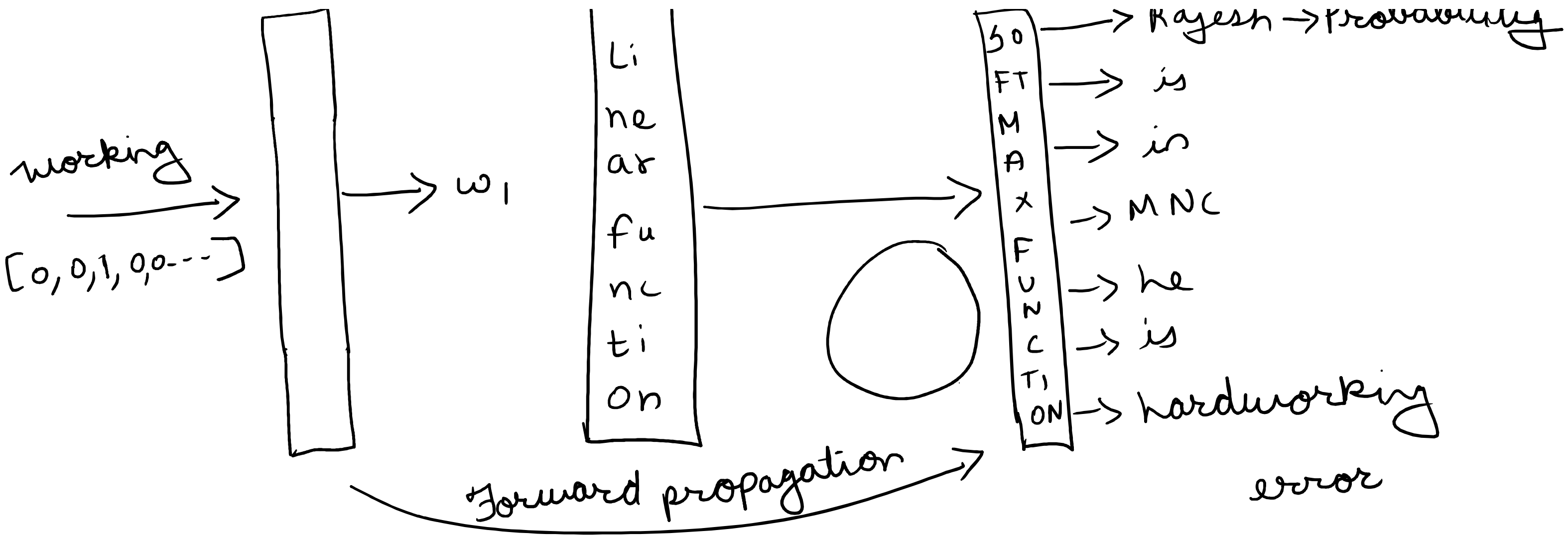
is = [0, 1, 0, 0, ---]

in = [0, 0, 0, 1, 0, ----]



Skipgram : →





Vector Representation of word = Weight Matrix × Probability of word.

Information extraction : Big umbrella

→ we are extracting knowledge from the text.

⇓

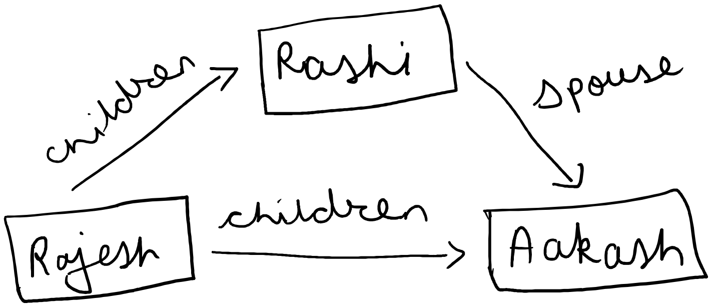
Sub-blocks

- Keyphrase extraction
- Named entity Recognition / extraction
- Relation extraction

Relation extraction: OpenNRE

Rajesh was born in 1940, the only child of teacher Aakash and his wife Rashi, a doctor...

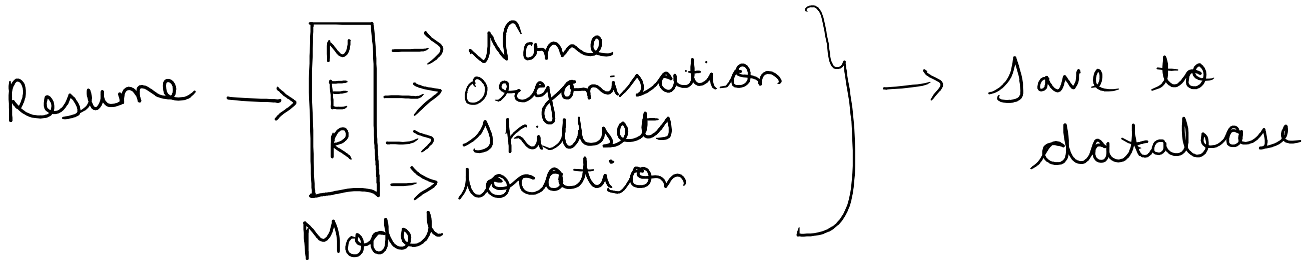
teacher Aakash and his wife Rashi, a doctor...



Named entity Recognition / extraction ⇒ Spacy

⇓

Resume Parsing



Keyphrase extraction :

-> extracting important phrases from the data.

Ways to do Keyphrase extraction

- ① Ngrams
- ② Unsupervised algorithms
  - RAKE, YAKE

Flow chart of unsupervised algorithm for Keyphrase extraction:

Keyphrase extraction.

Raw text ⟶ Candidate generation
↓
Candidate scoring

Final Ranking ⟵ Post processing ⟵ Candidate scoring

① Candidate generation

Extract all phrases

② Candidate scoring

give scores to phrases

Important phrases

③ Post processing ⇒ "game" ⟶ "games"

insert "s"

SUN ⟶ RAN

S -> delete

R -> insert ⇒ RAN

U -> delete

A -> insert

Laveshtein distance -> game -> games -> 1

L.D. -> 1

SUN -> RAN -> 4

L.D → 4

Use of Laveshtein distance:

Single Review → 1000 – 5000 words

Keyphrases → 2500

10 Documents → 1,00,000 – 5,00,000 words

Keyphrases → 2,50,000

Important Keyphrases → Keyphrases count ↓

Nice Movie]→ 2
Nicely Movie

⇓

Final Ranking: → 5,00,000 Phrases

⇓

1,50,000 Phrases

⇓

Top 100, 500

Applications of Keyphrase extraction: →

① Text summarization (extractive)

② Data analysis

③ Root cause analysis

④ Chatbot

$\underbrace{\text{Hi,}}$ How are you? ⇒ Hello, How are you?