SYNSET EXPANSION

106065502 何佳芳

WORD TO VECTOR
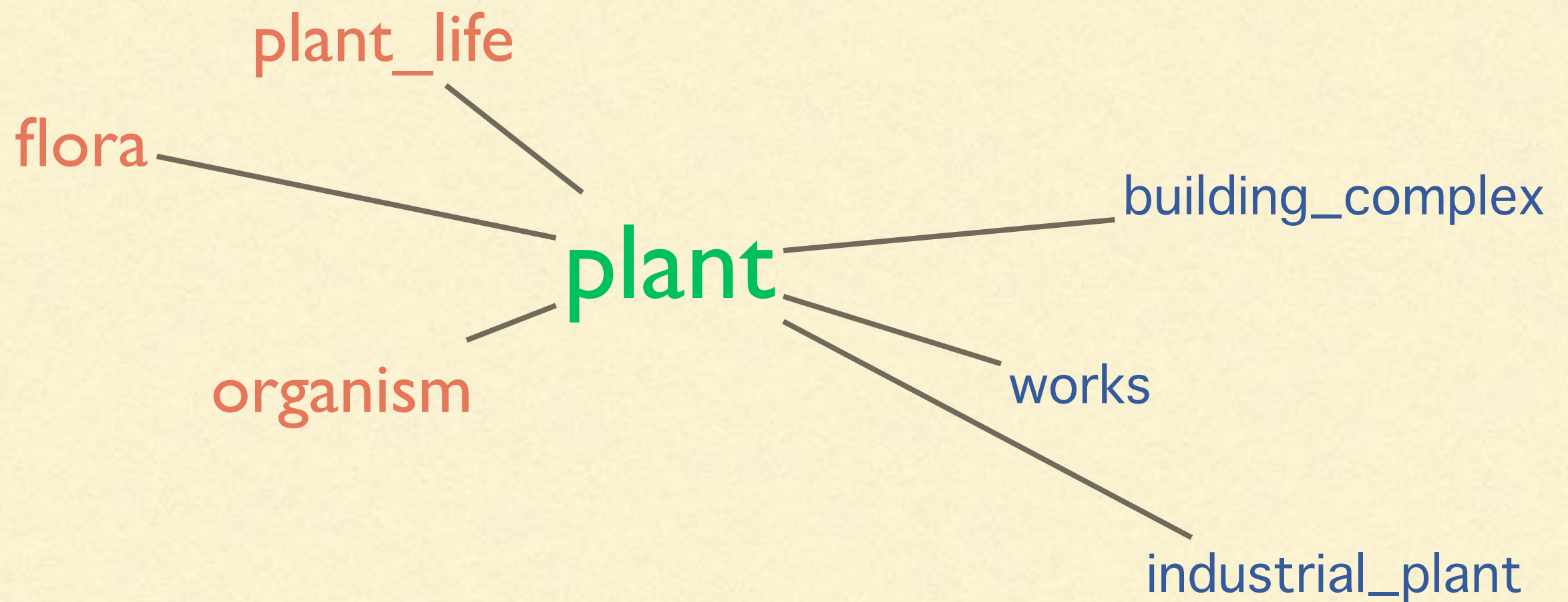
DATA

- WordNet

- Googlenews word2vec model

WN - PLANT

- S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
  - S: (n) living thing, animate thing (a living (or once living) entity)
    - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
      - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
        - S: (n) physical entity (an entity that has physical existence)
          - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

WORD2VEC most_similar()

- Googlenews

```
>>> model.most_similar(positive=['woman','king'],negative=['man']

>>>[('queen', 0.50882536), ...]
```

SYNSET

- [**plant**(organism)] = [plant_life] + [flora] + [organism] …

同義詞集

GET SYNSET

Get Word Sense

- Hypernyms

- Lemmas

Example: plant

Hypernyms	Lemmas	Definition
building_complex (建築物)	industrial_plant plant works	buildings for carrying on industrial labor
organism (植物)	plant_life flora plant	(botany) a living organism lacking the power of locomotion
actor (演員)	actor plant	an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience
contrivance (詭計)	plant contrivance	something planted secretly for discovery by another

Googlenews Word2Vec

- Positive = [ALL_SYNSET] / len(all_synset)

- Negative = [OTHER_SENSE] / len(other_sense)

Example: plant

Sense: (botany) a living organism lacking the power of locomotion

positive = [plant_life, flora, plant, organism]/4

negative = [plant, actor, contrivance]/3

```
>>> w2v.most_similars(positive, negative, topn = 200)
```

Filter

- Synonym in WordNet

- Similarity(synonym, word) > 0.2

Ranking

- Similarity(SYNONYM, WORD)

- std(frequency(SYNONYM))

DEMO

FUTURE

- Chinese sense

- Example

Q&A