

Corpus of texts

- Unstructured texts
- Semi-structured texts
e.g. wikipedia

Pre-processing

- Lemmatization
- Remove articles
- Remove stop words

...

Corpus of texts pre-processed

Corpora statistics

- word count
- tf_idf

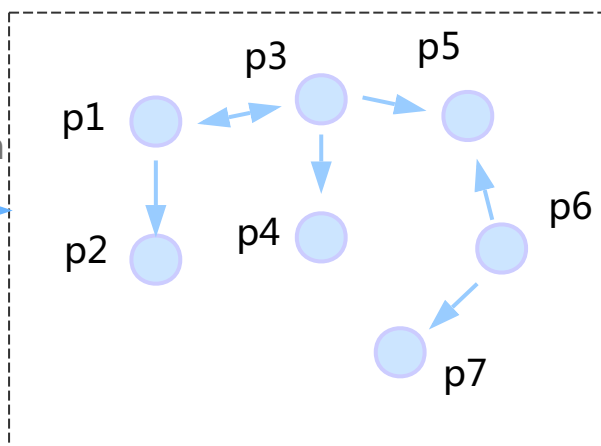
...

Vocabulary

w1
w2
w3
...

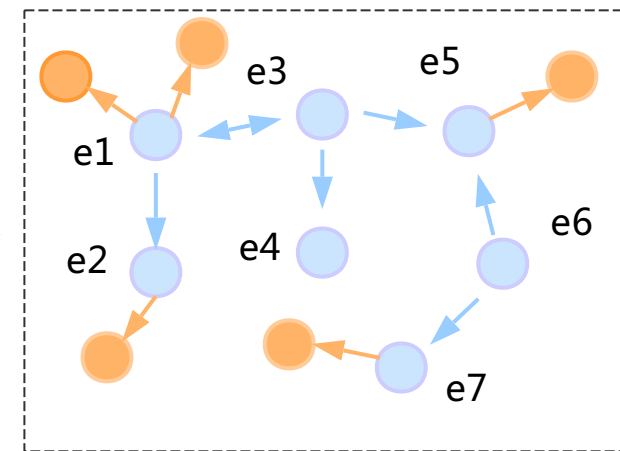
+

Information
Extraction



SPARQL

$$p_i = e_i \\ (2)$$



Transition Weight

Semantic Relatedness
Measurement

(3)

