

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

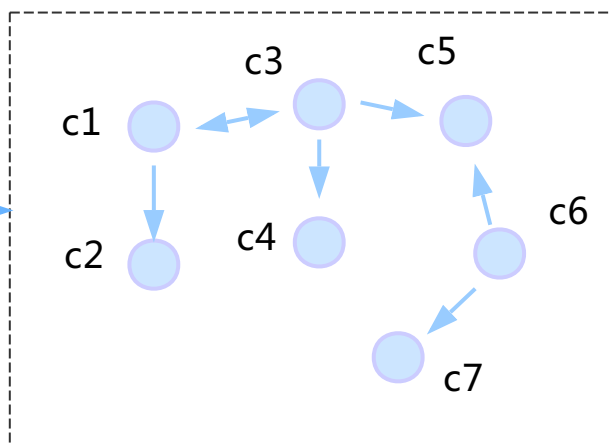
- word count
- PMI
- tf_idf
- ...

Vocabulary

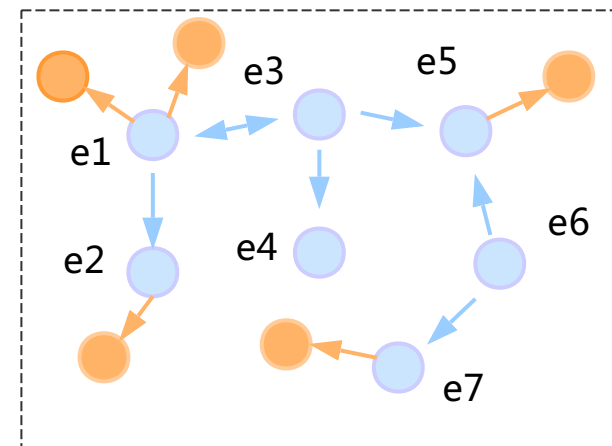
w1
w2
w3
...

+

Information
Extraction



SPARQL



Semantic Relatedness Measurement

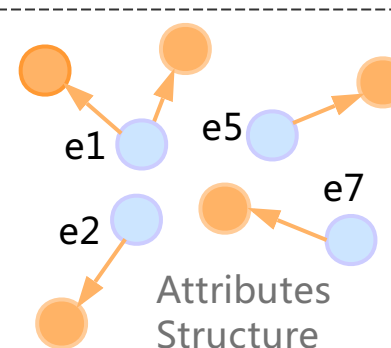
Weighting

c1 c2 c3 c4 c5 ...

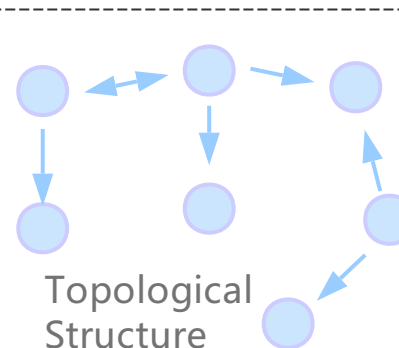
w1					
w2					
w3					
w4					
w5					
...					

Mapping between
words and concepts

Word-to-Concept



Attributes
Structure



Topological
Structure

Entity-to-Entity

$\vec{w_1}$

$\vec{w_2}$

Word-to-Word