NLP using Python

Stemming and Lemmatization

Stemming

"John does his work intelligently"

"John is an intelligent man"

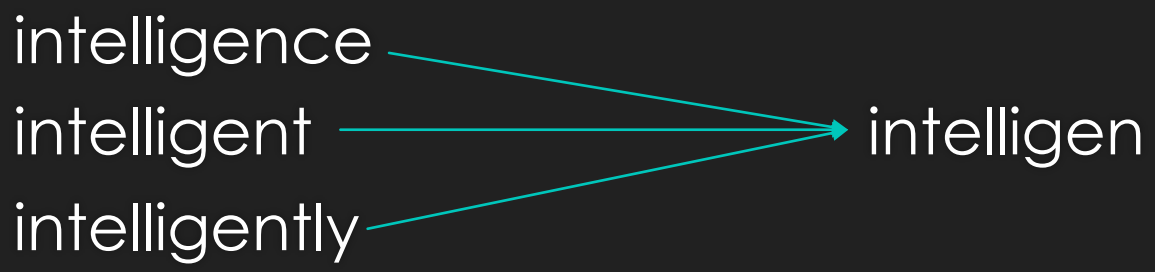
"John is always working"

Stemming

Sentence 1	Sentence 2	Sentence 3
John	John	John
does	is	is
his	an	always
work	intelligent	working
intelligently	man	

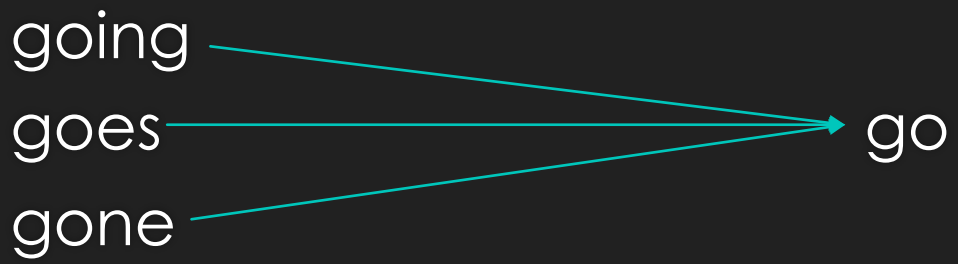
Stemming

"Stemming is process of reducing infected or derived words to their word stem, base or root form" – Wikipedia



Stemming

"Stemming is process of reducing infected or derived words to their word stem, base or root form" – Wikipedia



Stemming

Problem:

Produced intermediate representation of the word may not have any meaning.

Example: intelligen, fina etc

Lemmatization

"John does his work intelligently"

"John is an intelligent man"

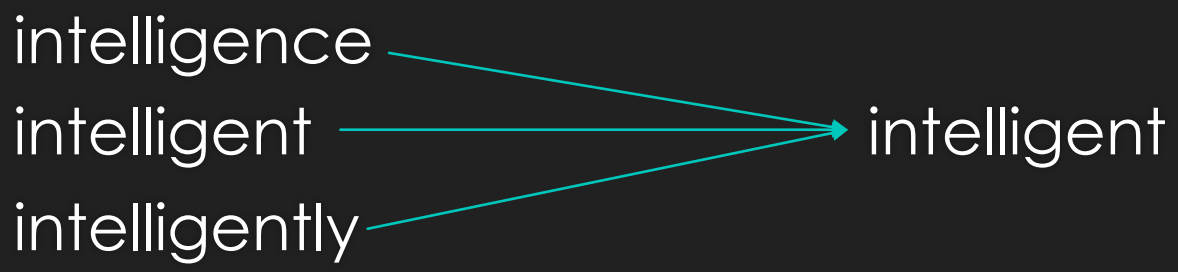
"John is always working"

Lemmatization

Sentence 1	Sentence 2	Sentence 3
John	John	John
does	is	is
his	an	always
work	intelligent	working
intelligently	man	

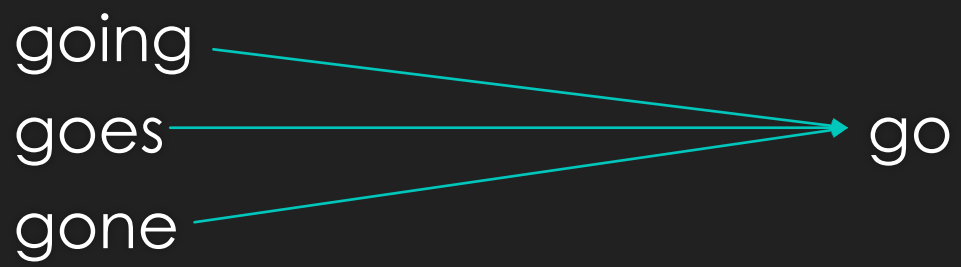
Lemmatization

Same as Stemming but intermediate representation/root form has a meaning



Lemmatization

Same as Stemming but intermediate representation/root form has a meaning



Lemmatization vs Stemming

Lemmatization

- Word representations have meaning.
- Takes more time than Stemming.
- Use lemmatization when meaning of words is important for analysis. Example: question answering application.

Stemming

- Word representations may not have any meaning.
- Takes less time.
- Use stemming when meaning of words is not important for analysis. Example: Spam detection.

Additional Read

Stemming and Lemmatization – Stanford NLP

https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html