Natural language Processing

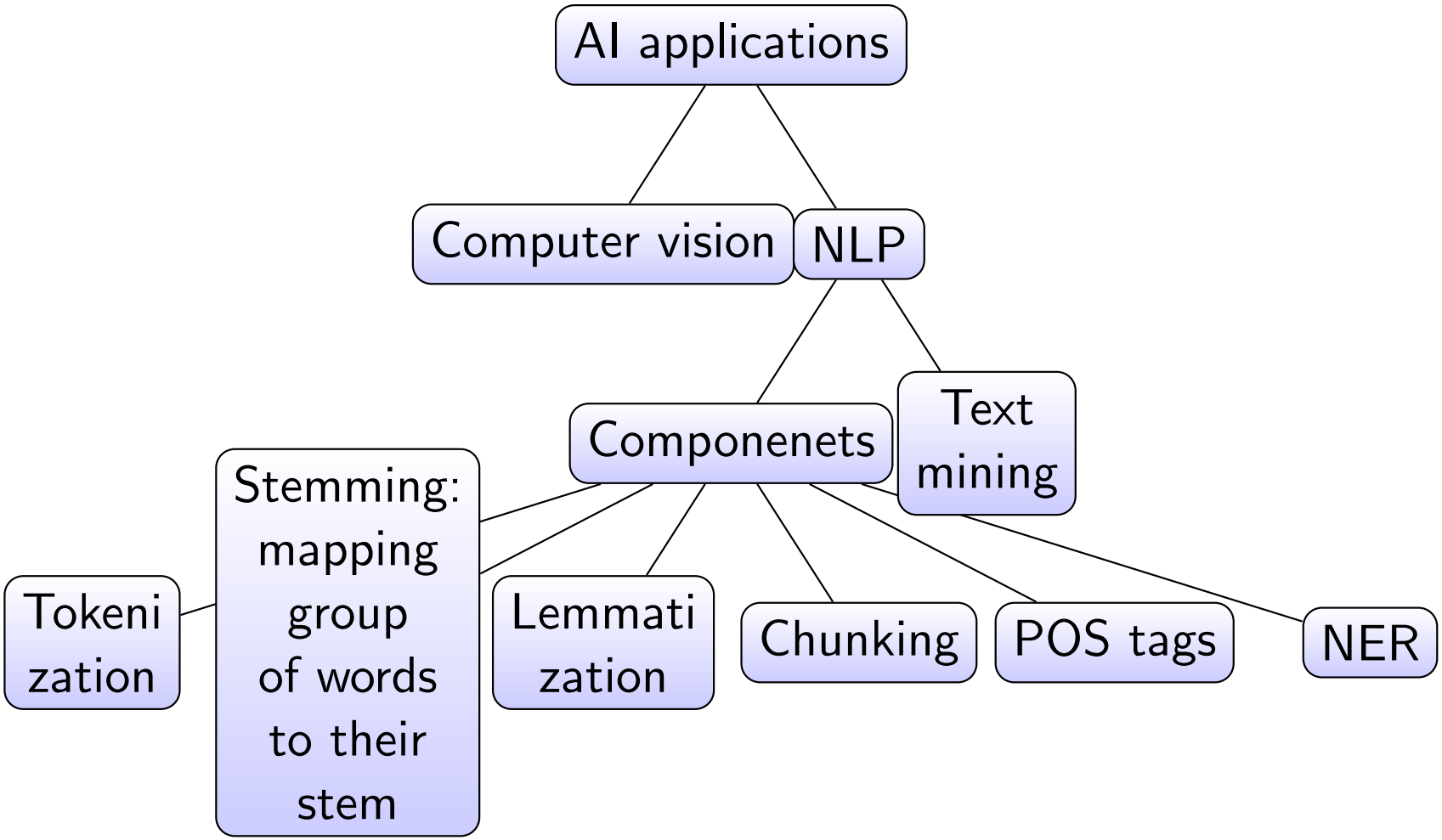
Saumya Bhatnagar

March 5, 2020

Table of contents I

General

Text Mining: process of deriving high quality information from texts
Turn text into data for analysis via NLP
NLP (AI): deals with human languages
what is inflection?


AI applications
Computer vision
NLP
Componenets
Text mining
Stemming: mapping group of words to their stem
Tokeni zation
Lemmati zation
Chunking
POS tags
NER

Tokenizer

break sentences into words

Stemming and Lemmatization

Stemming	**Lemmatization**
Might not be an actual word	Returns an actual word
Predefine steps	Uses WordNet corpus
Speed is imp	Language is imp

Stemming: Stem word might not be an actual word

- Porter stemmer: based on 5 pre-defined rules, simple and fast, mainly used in IR (information retreival) search queries
- Snow-ball stemmers: by NLTK, has non-english stemmers (french, german, english, italian,)
- Lancaster stemmer: Iterative, over-stemming may occur, shorter stem then Porter
- ISRI Stemmer
- RSLPS Stemmer

bayesian Modeling

content...

content...

content...

Thank You!