



Introduction to NLP

Video 5: Methods of Text Preprocessing - Part 2

Stemming

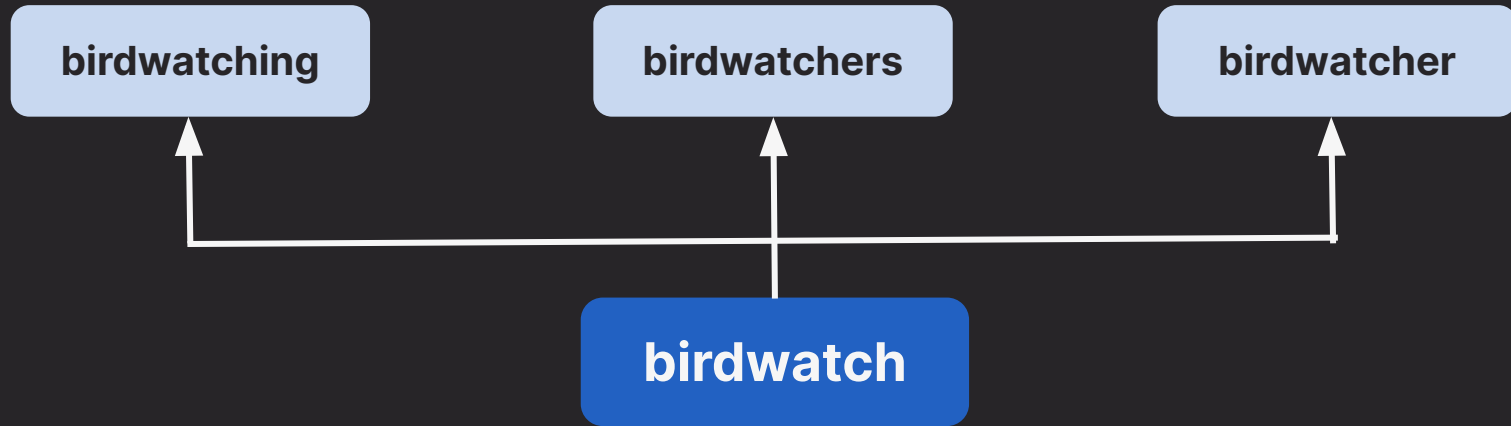
['birdwatching', 'serene', 'educational', 'hobby', 'captivated', 'enthusiasts', 'worldwide', 'offering', 'intimate', 'glimps e', 'avian', 'world', 'activity', 'deepens', 'one', 'appreciation', 'nature', 'diversity', 'also', 'enhances', 'observational', 'skills', 'birdwatchers', 'learn', 'identify', 'species', 'plumage', 'calls', 'behavior', 'process', 'begins', 'birdwatcher', 'positioning', 'habitats', 'ranging', 'dense', 'forests', 'urban', 'parks', 'armed', 'essential', 'tools', 'binoculars', 'fiel d', 'guides', 'guides', 'invaluable', 'distinguishing', 'species', 'noting', 'characteristics', 'color', 'patterns', 'sizes', 'distinctive', 'markings', 'additionally', 'listening', 'plays', 'crucial', 'role', 'birdwatching', 'ability', 'recognize', 'bi rdsongs', 'calls', 'reveal', 'presence', 'species', 'hidden', 'view', 'practice', 'also', 'involves', 'recording', 'observation s', 'often', 'journal', 'digital', 'app', 'contributes', 'citizen', 'science', 'providing', 'data', 'bird', 'populations', 'mig ration', 'patterns', 'aspect', 'birdwatching', 'underscores', 'significance', 'beyond', 'recreation', 'highlighting', 'contribu tion', 'conservation', 'efforts', 'endeavors', 'birdwatching', 'fosters', 'connection', 'nature', 'encourages', 'environmenta l', 'stewardship', 'provides', 'insights', 'ecological', 'roles', 'birds', 'within', 'habitats', 'serves', 'multifaceted', 'act ivity', 'combines', 'leisure', 'learning', 'conservation']

birdwatching

birdwatchers

birdwatcher

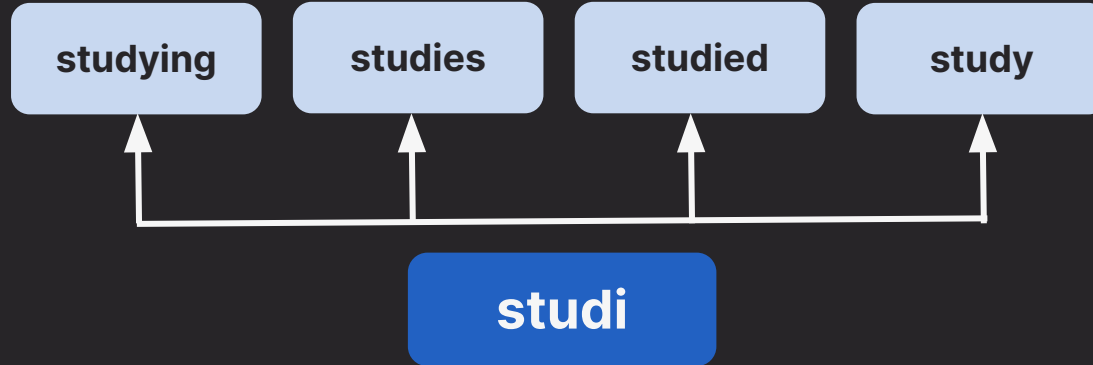
Stemming



Morphological base or root form

Stemming

Stemming reduces different variant of same word into its only root form.



Stemming



Removes suffixes from words.

-ing

-ed

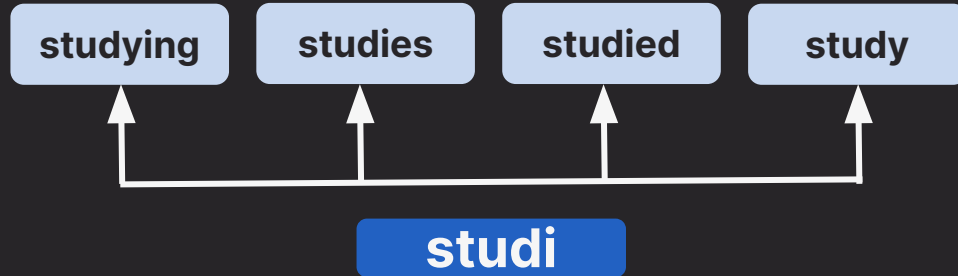
-ly

Stemming



Removes suffixes from words.

Doesn't consider the word's context or actual meaning.



Jupyter Notebook

Lemmatization

Lemmatization consider word's context and dictionary meaning.

Uses dictionary and morphological analyzer to identify lemma

Lemmatization

