Stemming

Stemming is the process of transforming words to their root form, even if the stem itself is not a valid word in the language

Staying, Stays, Stayed ⟶ Stay ← stem
house, houses, housing ⟶ hous

Stemming is a text preprocessing technique in NLP that reduces words to their base or root form

↳ The main purpose of stemming is to standardize words and reduce the vocabulary size which can improve the efficiency and effectiveness of various NLP tasks

running ⟶ run
fishes ⟶ fish
argued ⟶ argu

cats, catlike, catty ⟶ cat
friendship, friendly, friends ⟶ friend
universe, university, universal ⟶ values

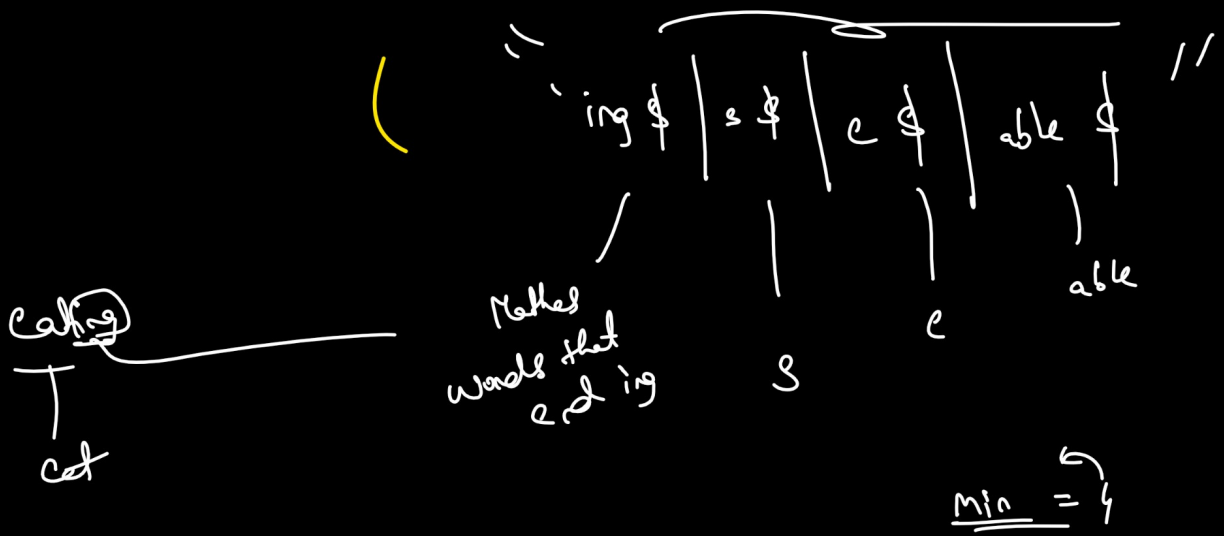
A root word is the most basic part of a word ⟶ core meaning without any prefixes or suffixes

Types of Stemming

1. Porter stemmer
2. Snowball stemmer ⟶ "language"

Applications

1. Information retrieval
2. text classification
3. Sentiment Analysis
4. topic modeling //



Lemmatization ?

Lemmatization is quite similar to stemming but unlike stemming . it reduces the words to looks that are valid Words in the Language

Stay, stays, staying, stayed → stay } lemma

house, houses, housing → house

text Normalization techique used in NLP to reduce words to their base or dictionary form known as lemma

running → run

better → good

corpora → corpus

1. chatbots and virtual assistants
2. Text classification & clustering
3. Sentiment Analysis
4. Big data Analytics

Stemming

1. produces roots of words
2. Fast and efficient to compute

Lemmatization

1. produces actual words
2. slower than stemming and can depend on the part of speech