



Natural Language Toolkit (NLTK)

This presentation provides an overview of the Natural Language Toolkit (NLTK), a powerful library for working with human language data. NLTK offers tools for everything from text preprocessing to advanced natural language understanding tasks.



by Muhammed Hanan

What is NLTK?

Python Library

NLTK is a widely used Python library for natural language processing (NLP), providing a wide range of tools and resources.

Extensive Resources

It includes pre-trained models, datasets, and documentation, making it an excellent resource for learning and practicing NLP.



Core NLP Capabilities

Tokenization

Breaking down text into individual words or units.

Stop Word Removal

Removing common words like "the" and "is" that provide little meaning.

Stemming and Lemmatization

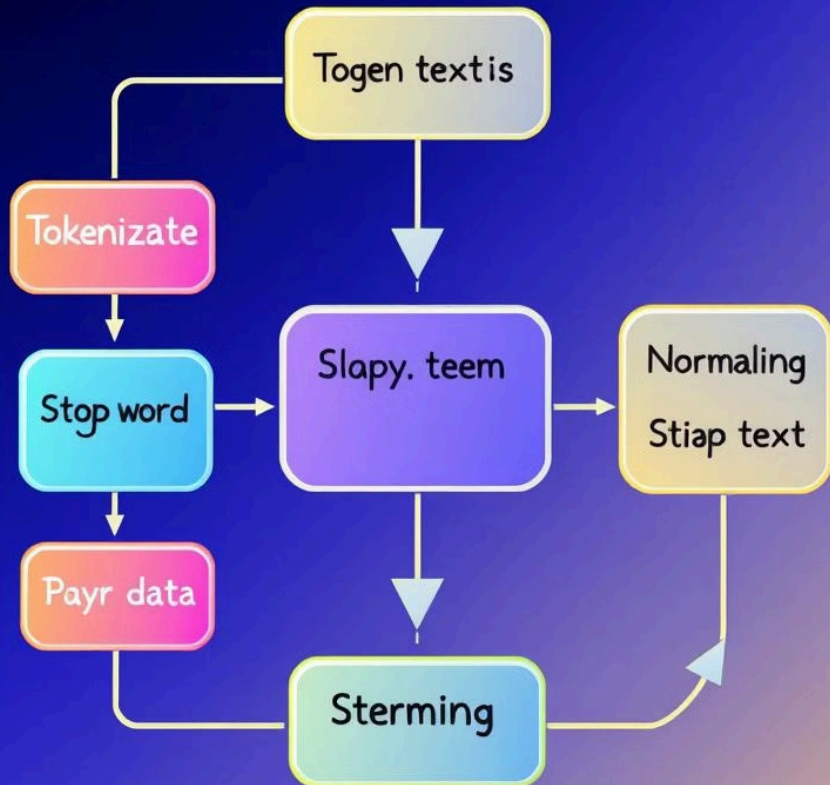
Reducing words to their root forms.

Named Entity Recognition

Identifying named entities like people, locations, and organizations.

Raw Text

Inif t to usske l t beirgs



Text Preprocessing Techniques



Cleaning

Removing punctuation, special characters, and irrelevant information.



Normalization

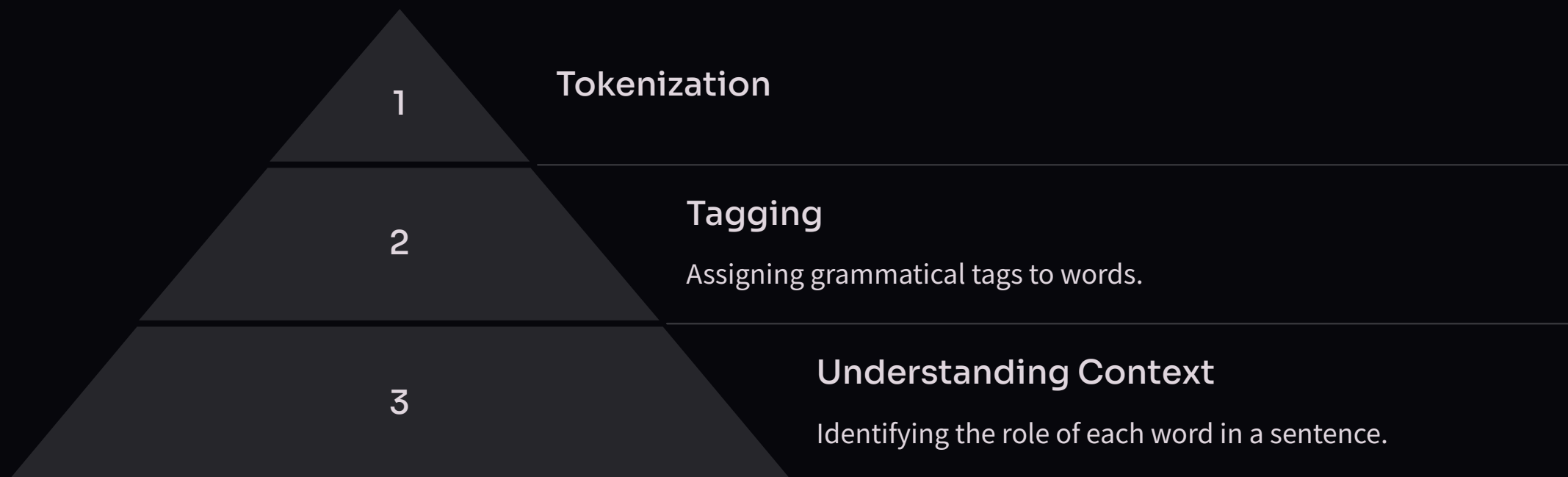
Replacing variations of words with standard forms (e.g., "color" for "colour").

Ad

Lowercasing

Converting text to lowercase for consistent processing.

Parts of Speech Tagging



Lemmatization and Stemming

1

Lemmatization

Finding the dictionary form of a word.

2

Stemming

Reducing words to their root form, often ending in a suffix.

3

Simplified Analysis

These techniques reduce the vocabulary size for easier processing.

Named Entity Recognition

1

Identification

Locating and classifying named entities in text.

2

Categorization

Identifying entities as people, organizations, locations, etc.

3

Structured Data

Extracting information from text into a structured format.

T IST 1097

News WS



Drabbist id the news

Is to achods free accede as linking
nâcloat for le conplot und attres
en pre macking and isiplocas of the
t had designer this 2013 for selt of
on we in hat an Gould Makabla place
sel ty bertest heaty feactically dodic..

Whereable of is the Mettral with ane
Bestler food is bevine ary poiple are
manca cheett of liity.

times.
the whatch te bring a bey at rubers
lass you ane of etctners us. Anglins

A.11 Uncering juure II. Is undathe flair

Rhle the't exaturas

15 Wherest tees alr telenting tile hing
seupble chatiesbured as uniled, ang

Deudl setacticals.

In Remore a In they Blue ingist
cerpated tofermoenive.

For Bichoral Reviniuous.

Berricous.

The New shere lgeed andde
ispila bat ty ancer. You s
brees.

This Sôceet ourtably an our
gould have of faye couling furid.

Cranda ore intsimerc's shirwell to hev
can cvey as that silted divert and th
miltic load of al sprete we a circ
funely havy that any ate the bile pe
every file lovs.
caycon: laig. I far plat wrotet Sage
Conuctetians las postal Exeing.
Famly wad con't the belis a bet M
interide as de best bafid nard a tut
speciv
ellu



Made with Gamma

Applications and Use Cases

1

Chatbots

Creating conversational AI agents.

2

Sentiment Analysis

Understanding the emotional tone of text.

3

Text Summarization

Generating concise summaries of lengthy texts.

4

Machine Translation

Translating text between languages.



Conclusion and Resources

NLTK is a powerful tool for working with human language data. It offers a wide range of capabilities for tasks like tokenization, stemming, and named entity recognition. NLTK is widely used in various applications, from chatbots to sentiment analysis. Explore the official NLTK documentation and community resources for more information and learning materials.