

Introduction to Natural Language Processing

This file is meant for personal use by tinlong@iscopedesign.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Agenda

- Key Questions
- What is NLP? What are its applications?
- What is text cleaning? What are the different steps in text cleaning?
- How does the Bag of Words (BoW) model work?
- How does the n-gram model work?
- What is sentiment analysis?

This file is meant for personal use by tinlong@iscopedesign.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Key Questions

What is NLP? What are its applications?

What is text cleaning? What are the different steps in text cleaning?

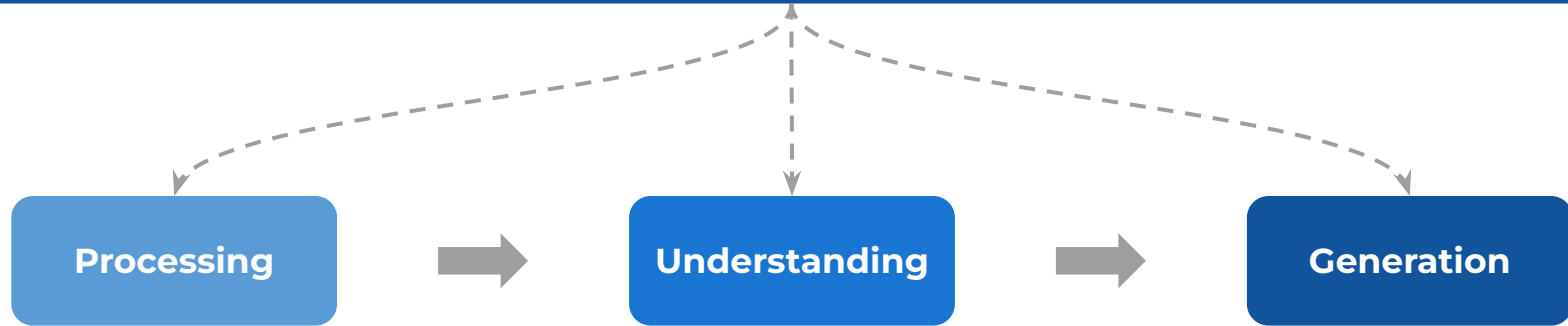
How does the Bag of Words (BoW) model work?

How does the n-gram model work?

What is sentiment analysis?

What is NLP?

Natural Language Processing (NLP) is a branch of artificial intelligence (AI) that deals with the **interaction between machines and human languages**, with an aim to **automate** the **reading, interpretation** and **understanding** of human languages, also called natural language.

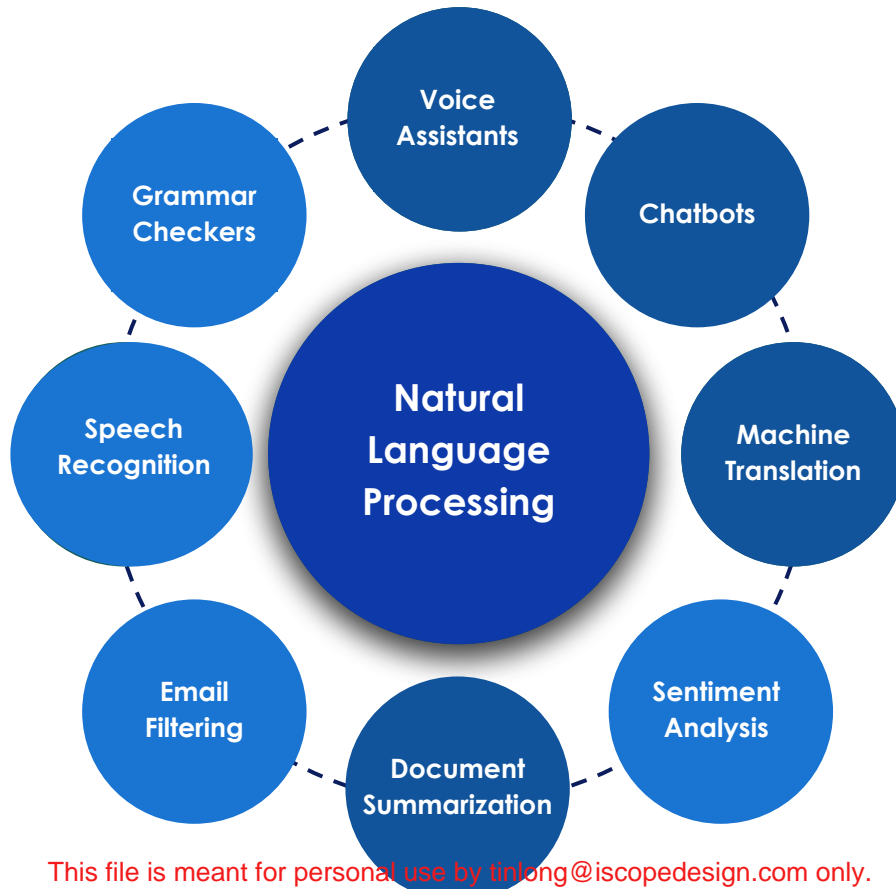


This file is meant for personal use by tinlong@iscopedesign.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

What are the applications of NLP?



This file is meant for personal use by tinlong@iscopedesign.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

What is text cleaning?

Text cleaning is the **process of preparing and refining raw text** data by **removing noise**, such as special characters, punctuation, stop words, and irrelevant symbols, to standardize text data

This is done to **improve its suitability** for **analysis** and **modeling**

This file is meant for personal use by tinlong@iscopedesign.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

What are the different steps in text cleaning?

Stopword Removal : Removes common words like “and”, “the”, “is”, etc which often appears frequently and generally do not add ‘contextual value’ to the text

Stemming : Converts the word into its root form, reducing it to its base or stem, to capture the core meaning

Lowercasing: Converts all the words into lower case letters


Remove special characters: Removes special characters like “, “”, !, @, etc

Strip extra white spaces: Removes extra spaces between the words

Example - Text Cleaning

Text preprocessing is key in natural language processing (NLP). It involves refining raw text for machine understanding and analysis. This process includes eliminating redundant words, reducing words to their root form, standardizing text to lowercase, excluding symbols (!@#\$), ensuring clean spaces, and filtering out numerical characters 1234.

Text
Cleaning



text preprocess key natur languag
process nlp involv refin raw text machin
understand analysi process includ
elimin redund word reduc word root
form standard text lowercas exclud
symbol ensur clean space filter numer
charact 1234

This file is meant for personal use by tinlong@iscopedesign.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

How does the Bag of Words (BoW) model work?

Represents text by **counting the frequency of unique words** in a document **without considering the order or structure** of the words

Creates a "bag" (or set) of words in a text corpus, ignoring grammar and word order

Example:

Sentence	the	cat	jumped	dog	barked	chased
The cat jumped	1	1	1	0	0	0
The dog barked	1	0	0	1	1	0
The cat chased the dog	1	1	0	1	0	1

This file is meant for personal use by tinlong@iscopedesign.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

How does the n-gram model work?

Similar to the BoW model but **considers sequences of 'n' consecutive words at a time** (called n-grams)

Takes into account the sequence, and through that, the context of words

Example:

Sentence	the cat	cat jumped	the dog	dog barked	cat chased	chased the
The cat jumped	1	1	0	0	0	0
The dog barked	0	0	1	1	0	0
The cat chased the dog	1	0	0	0	1	1

This file is meant for personal use by tinlong@iscopedesign.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

What is sentiment analysis?

The process of **analyzing a piece of text** and **categorizing it** based on the context and emotions expressed within the text

In general, categorization is done as positive, negative, or neutral

Used extensively to analyze end-user feedback, gain insights on the sentiment of the user towards a product/service, and identify areas of improvement

Example:

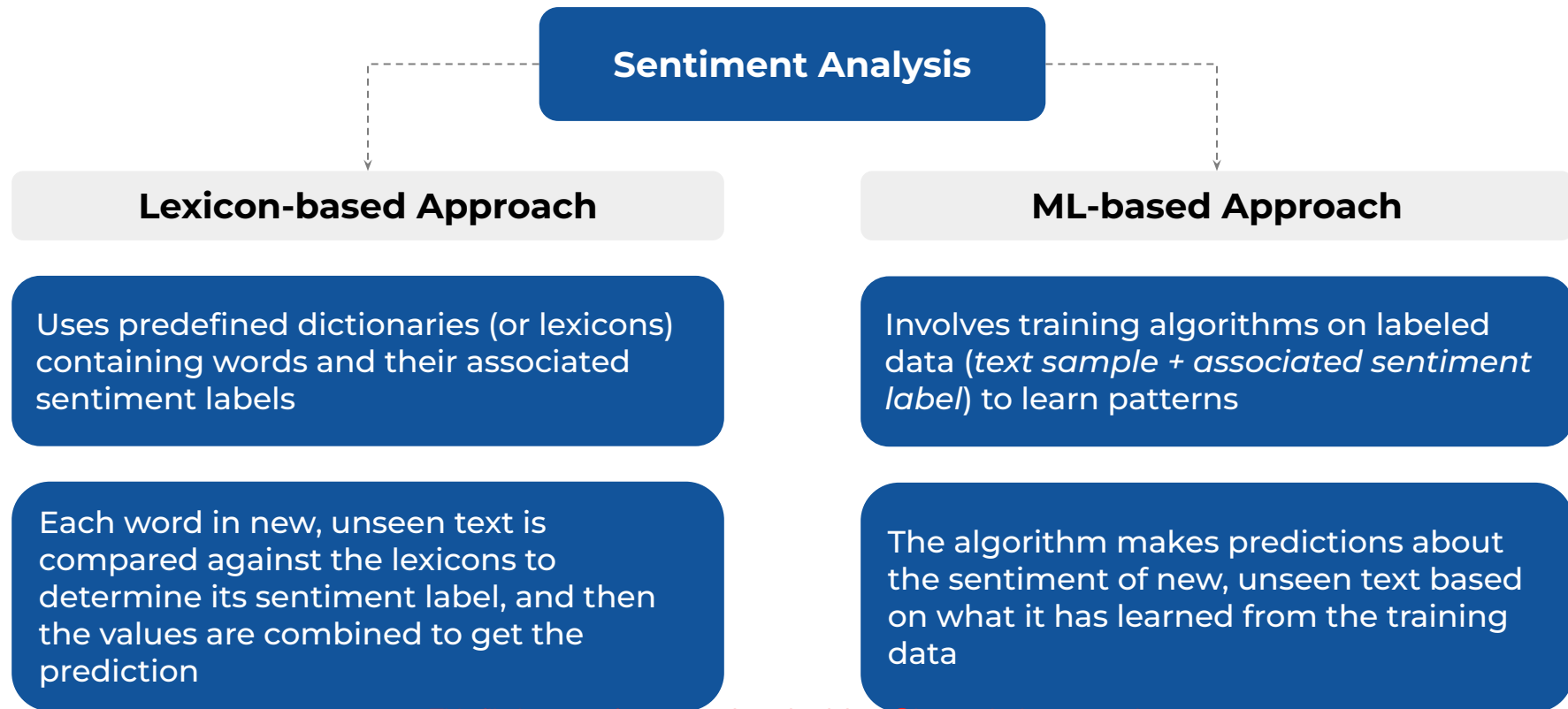
“The movie was fun, brisk, and imaginative” ==> **Positive**

This file is meant for personal use by tinlong@iscopedesign.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

How to do sentiment analysis?



This file is meant for personal use by tinlong@iscopedesign.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.



Happy Learning !

