simplilearn | PURDUE UNIVERSITY

Natural Language Processing

©Simplilearn. All rights reserved.

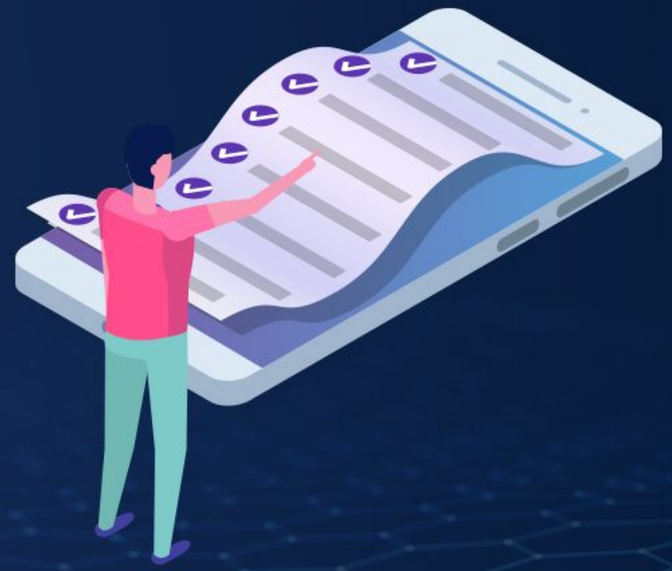
Natural Language Understanding Techniques

©Simplilearn. All rights reserved.

Learning Objectives

By the end of this lesson, you will be able to:

- Define parts-of-speech tagging
- Explain the different parsing methods
- Describe morphological techniques



©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE

Parts-of-Speech Tagging

©Simplilearn. All rights reserved.

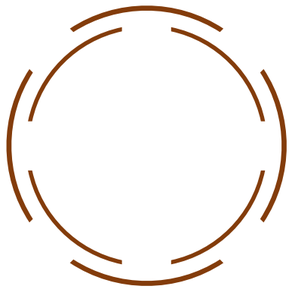
Parts-of-Speech Tagging

- It is used to understand the POS in a phrase or sentence.
- It assigns POS tag to each word in a phrase or sentence.
- POS tags include:
  - NN (Noun, singular)
  - VB (Verb, base form)
  - CD (Cardinal number)
  - RB (Adverb)
- A set of all POS tags used in a corpus is called a tagset.

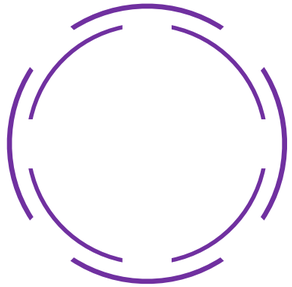
©Simplilearn. All rights reserved.

Parts-of-Speech Tagging

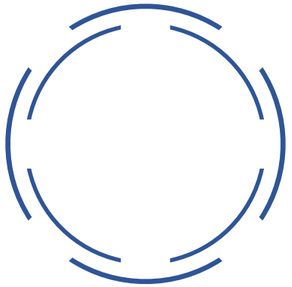
The parts-of-speech in the English language are:



Noun

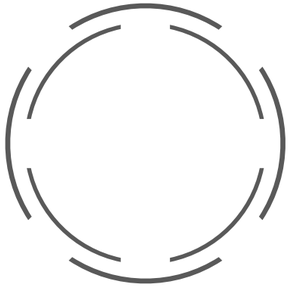


Pronoun

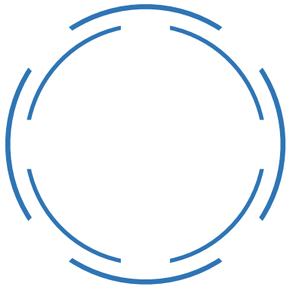


Verb

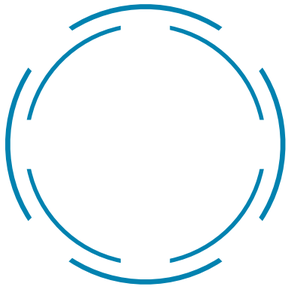
Adjective



Adverb



Preposition

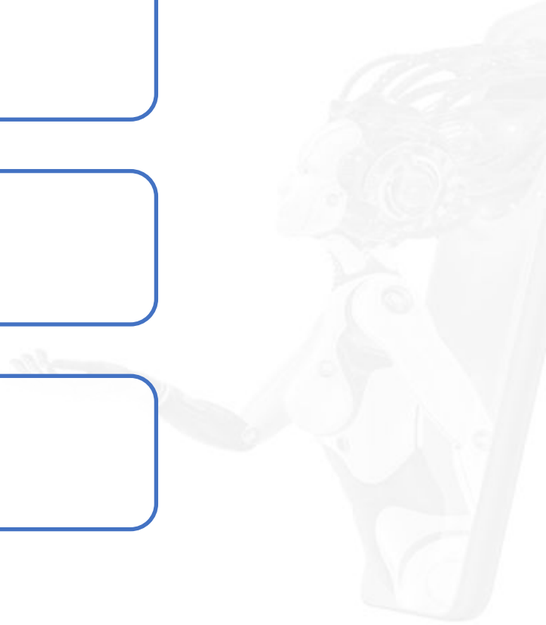


Conjunction

©Simplilearn. All rights reserved.

Parts-of-Speech Tagging: Advantages

1. Has better understanding of the context of the sentence
2. Identifies the relationship between words
3. Recognizes cross-references



©Simplilearn. All rights reserved.

simplilearn

Parts-of-Speech Tagging: Tags

Tag	Description	Example
NN	Noun, Singular	Chair
NNS	Noun, Plural	Chairs
NNP	Proper Noun, Singular	Mathew
PRP	Pronoun Personal	He
PRPS	Pronoun Possessive	His
VB	Verb, Base form	Run
JJ	Adjective	Good
JJR	Adjective Comparative	Better
JJS	Adjective Superlative	Best
RB	Adverb	Naturally
CC	Coordinating conjunction	And
DT	Determiner	The
CD	Cardinal number	I
POS	Possessive ending	Mathew's

©Simplilearn. All rights reserved.

Parts-of-Speech Tagging: Tags

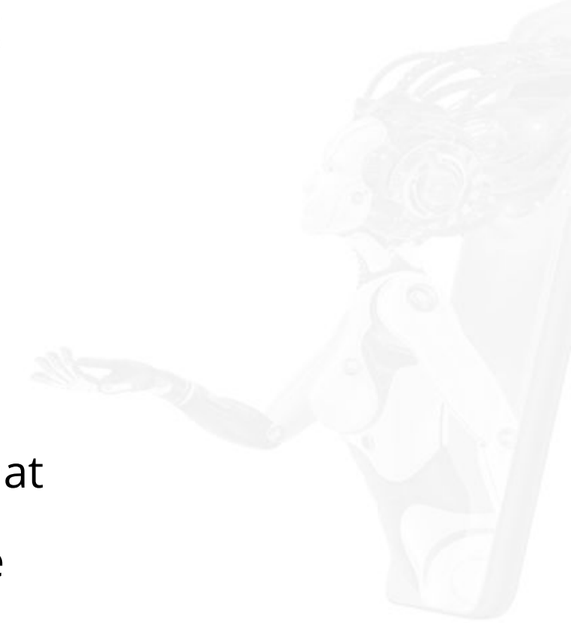
Tag	Description	Example
EX	Existential there	there is
FW	Foreign word	d'hoevre
IN	Preposition or subordinating conjunction	in, of, like
LS	List item marker	1)
MD	Modal	could, will
NNPS	Proper noun, plural	doors
PDT	Predeterminer	both the boys
RBR	Adverb, comparative	better
RBS	Adverb, superlative	best
RP	Particle	give up
SYM	Symbol	@, *
TO	*to*	to go, to him
UH	Interjection	uhhuhhuhh
VBD	Verb, past tense	took

©Simplilearn. All rights reserved.

simplilearn

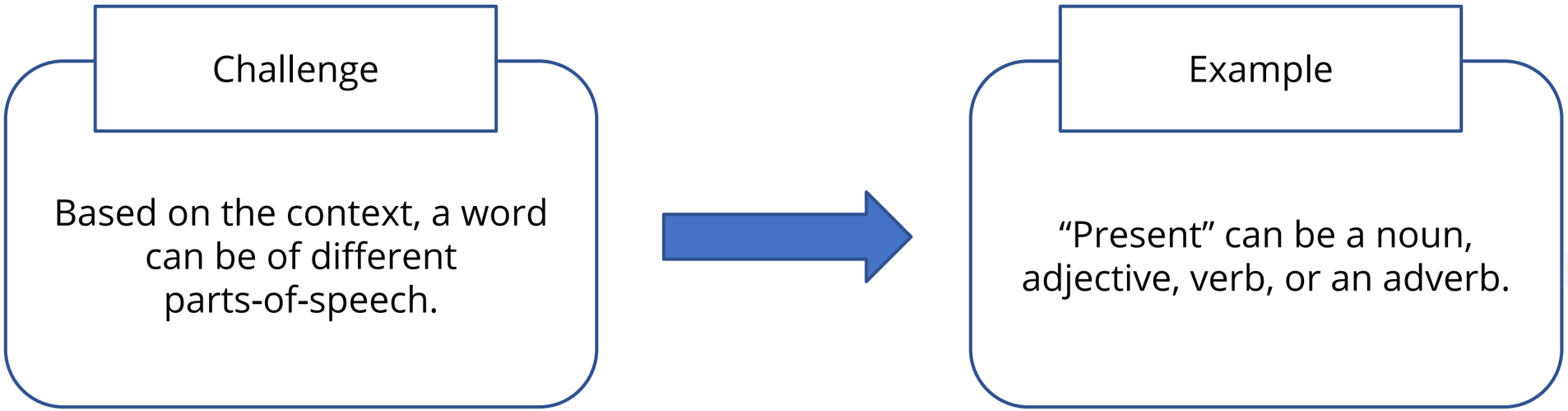
Parts-of-Speech Tagging: Tags

Tag	Description	Example
VBG	Verb, gerund or present participle	taking
VBN	Verb, past participle	taken
VBP	Verb, non-3rd person singular present	take
VBZ	Verb, 3rd person singular present	takes
WDT	Wh-determiner	which
WP	Wh-pronoun	who, what
WP$	Possessive wh-pronoun	whose
WRB	Wh-adverb	where, when



©Simplilearn. All rights reserved.

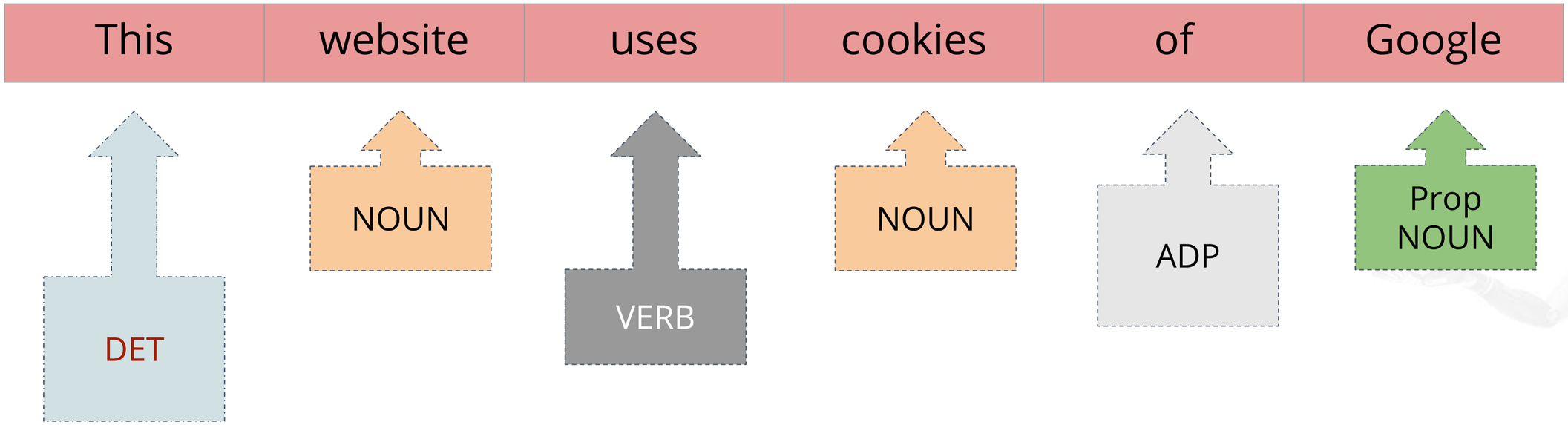
Parts-of-Speech Tagging: Challenges



©Simplilearn. All rights reserved.

Parts-of-Speech Tagging: Example

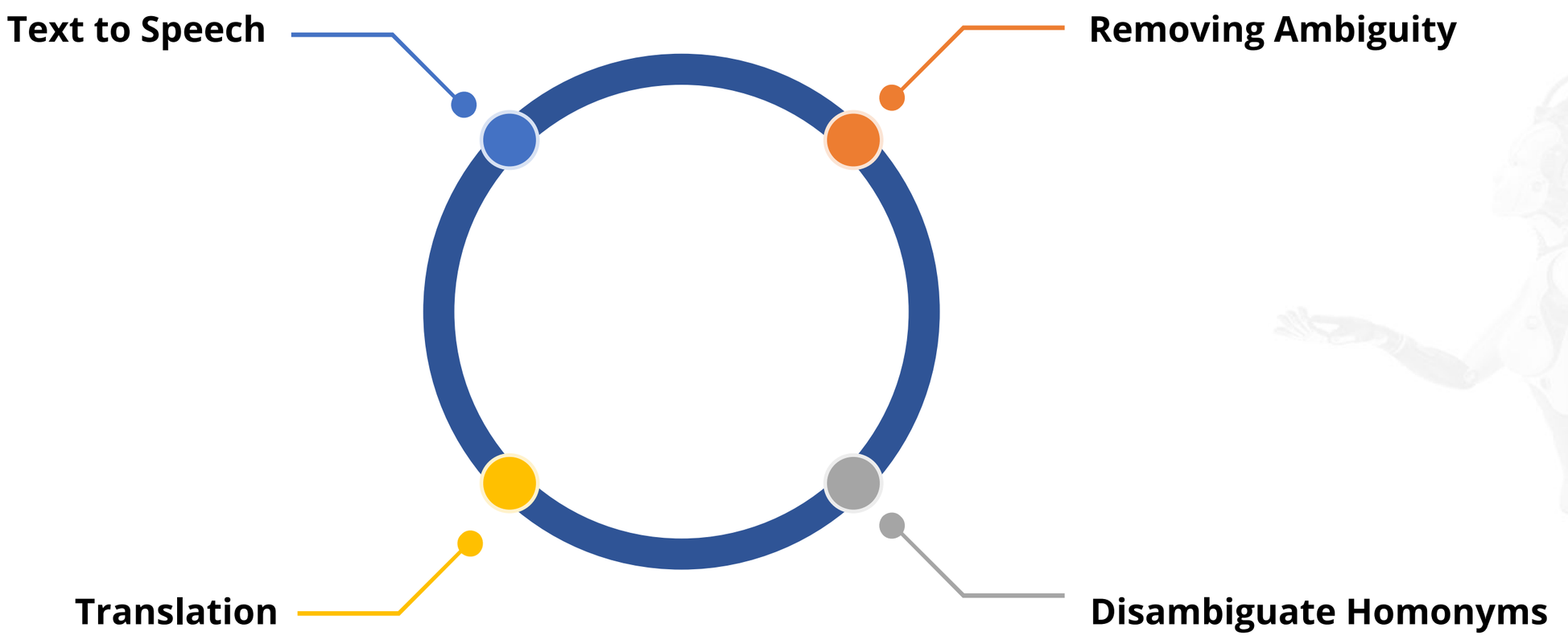
Sentence: This website uses cookies of Google.



©Simplilearn. All rights reserved.

simplilearn

Parts-of-Speech Tagging: Usage



©Simplilearn. All rights reserved.

Apply Fuzzy Search to Identify Similar Words

Problem Statement: Parts-of-speech plays a very crucial role in creating a grammatically correct sentence. To understand the sentence structure we can use POS tag identification techniques. Find a noun word from the given sentence and use this keyword to find similar documents.

Access: Click on the **Practice Labs** tab on the left side panel of the LMS. Copy or note the username and password that is generated. Click on the **Launch Lab** button. On the page that appears, enter the username and password in the respective fields, and click **Login**.

ASSISTED PRACTICE

©Simplilearn. All rights reserved.

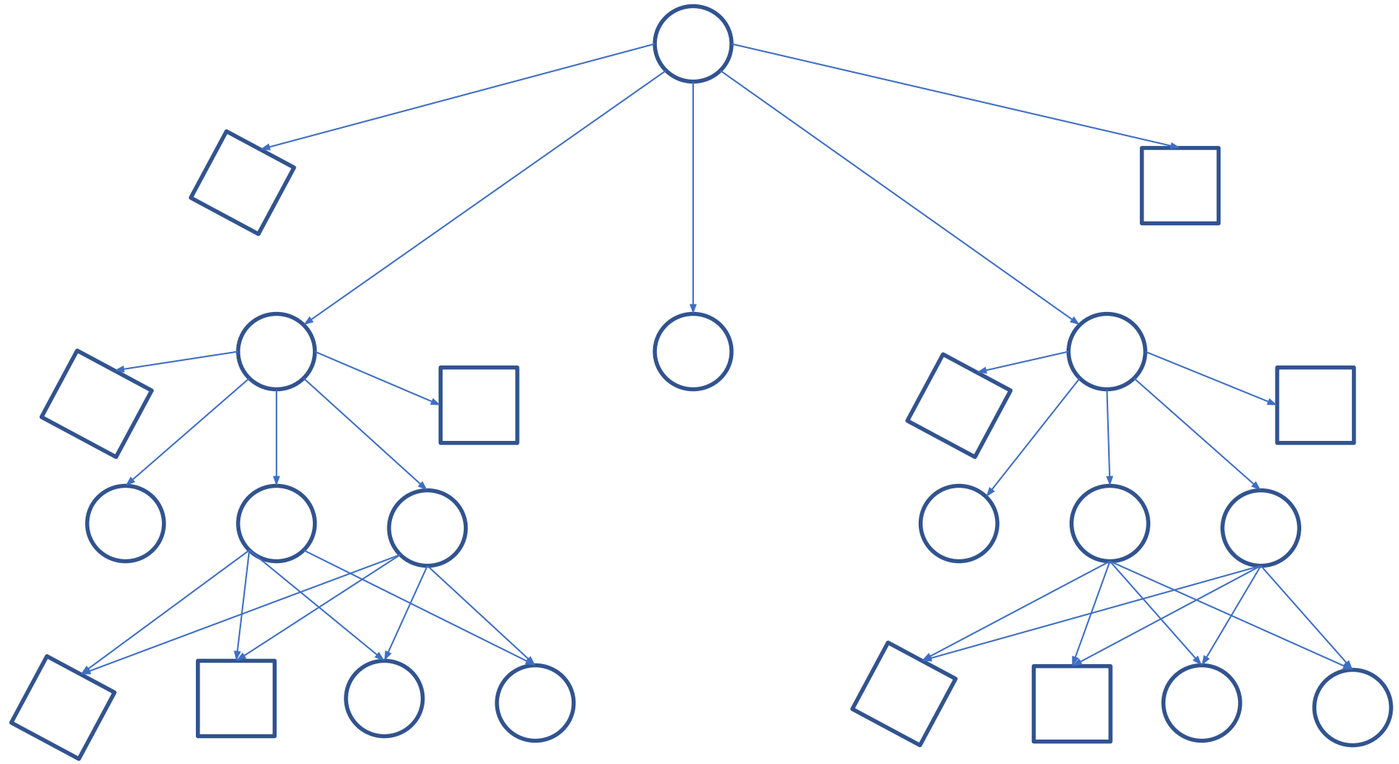
DATA AND
ARTIFICIAL INTELLIGENCE

Dependency Parsing

©Simplilearn. All rights reserved.

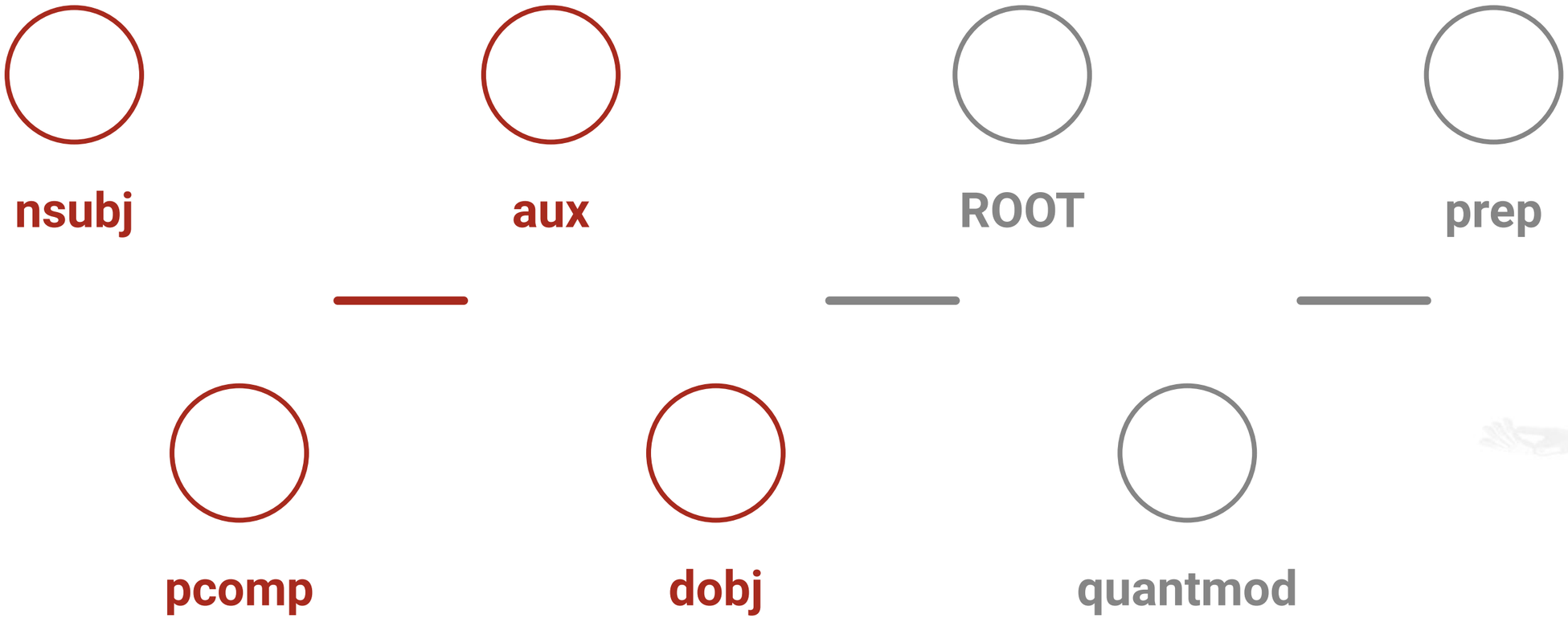
Dependency Parsing

- A dependency parser analyzes the grammatical structure of a sentence.
- It finds relationships between **head** words and words which modify those heads.
- Relations among the words are illustrated in a tree called the dependency tree.



©Simplilearn. All rights reserved.

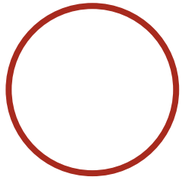
Dependency Parsing: Modifiers



©Simplilearn. All rights reserved.

simplilearn

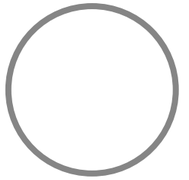
Dependency Parsing: Modifiers



nsubj

Auxiliary

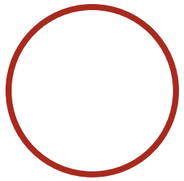
Expresses categories such as tense, aspect, and mood



ROOT

Nominal Subject

- Is the syntactic subject and the proto-agent of a clause
- Is only applied to semantic arguments of a predicate



aux

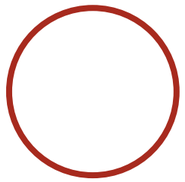
Root

- Is indexed with 0
- Is the only modifier with root dependency

©Simplilearn. All rights reserved.

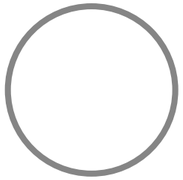
simplilearn

Dependency Parsing: Modifiers



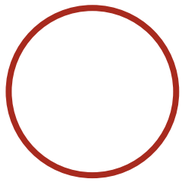
prep

Prepositional modifier



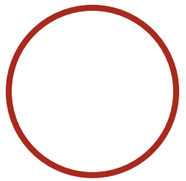
pcomp

Complement of preposition



dobj

Direct object



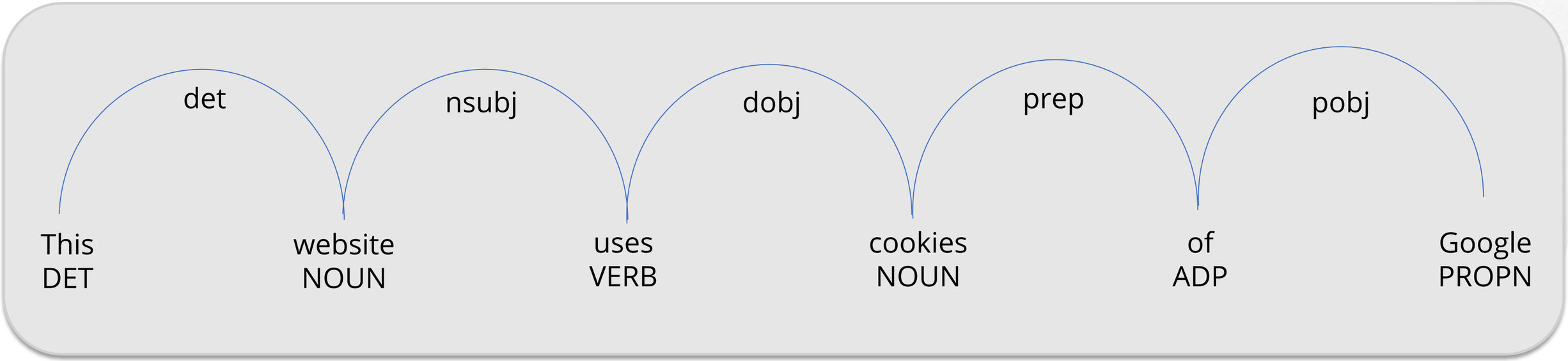
quantmod

Modifier of quantifier

©Simplilearn. All rights reserved.

Dependency Parsing: Example

Sentence: **This website uses cookies of Google.**



©Simplilearn. All rights reserved.

simplilearn

Dependency Parsing: Types

Grammar-Driven (Rule-based)

Data-Driven (Machine learning based)

©Simplilearn. All rights reserved.

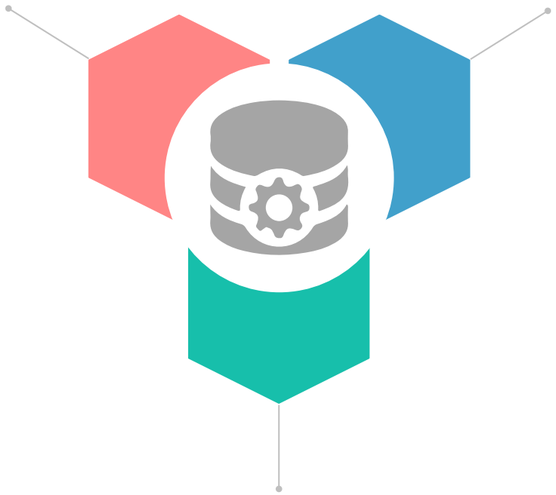
simplilearn

Dependency Parsing: Pros and Cons

Below are the pros and cons of dependency parsing:

Word Order

- Structure is independent of word order
- Suitable for free word order languages



Limited Expressivity

- One node per word
- Hard to find phrase modification and head modification in unlabeled dependency structure

Transparency

- Direct encoding of predicate-argument structure

©Simplilearn. All rights reserved.

Dependency Parsing: Usage

©Simplilearn. All rights reserved.

simplilearn

DATA AND
ARTIFICIAL INTELLIGENCE

Constituency Parsing

©Simplilearn. All rights reserved.

Constituency Parsing

1. It extracts a constituency-based parse tree from a sentence.
2. It denotes the syntactic structure according to phrase structure grammar.
3. Nonterminals are a type of phrases.
4. Terminals are words.

©Simplilearn. All rights reserved.

Constituency Parsing

Syntactic categories in NLP:

- NP: Noun Phrase
- VP: Verb Phrase
- S: Sentence
- DET: Determiner
- N: Noun
- TV: Transitive Verb
- IV: Intransitive Verb
- PRP: Preposition
- PP: Prepositional Phrase
- ADJ: Adjective

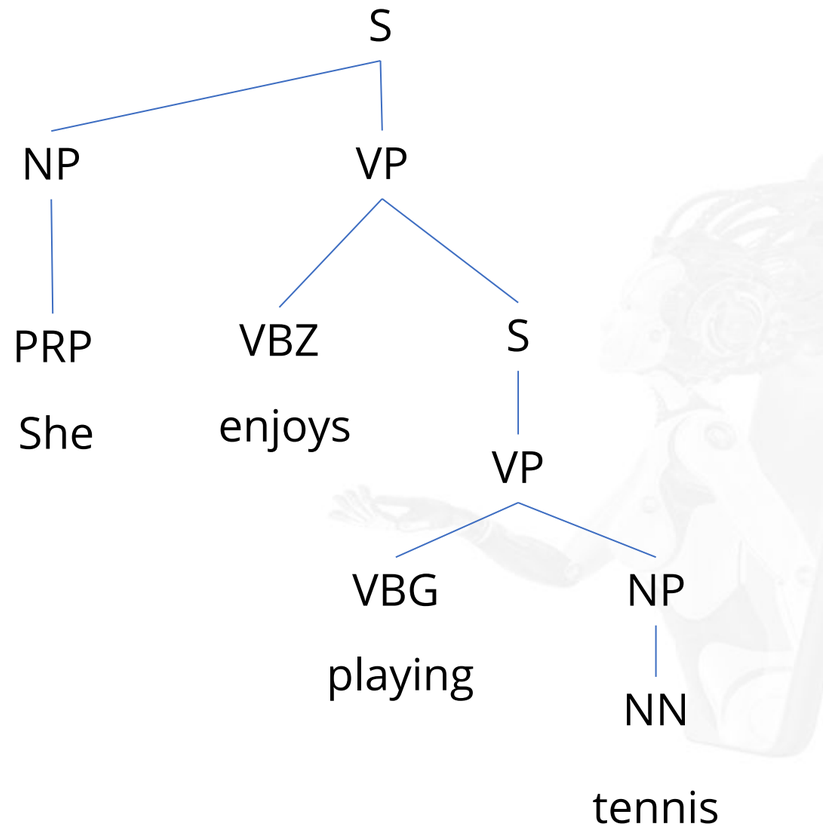
©Simplilearn. All rights reserved.

simplilearn

Constituency Parsing

Sentence : She enjoys playing tennis

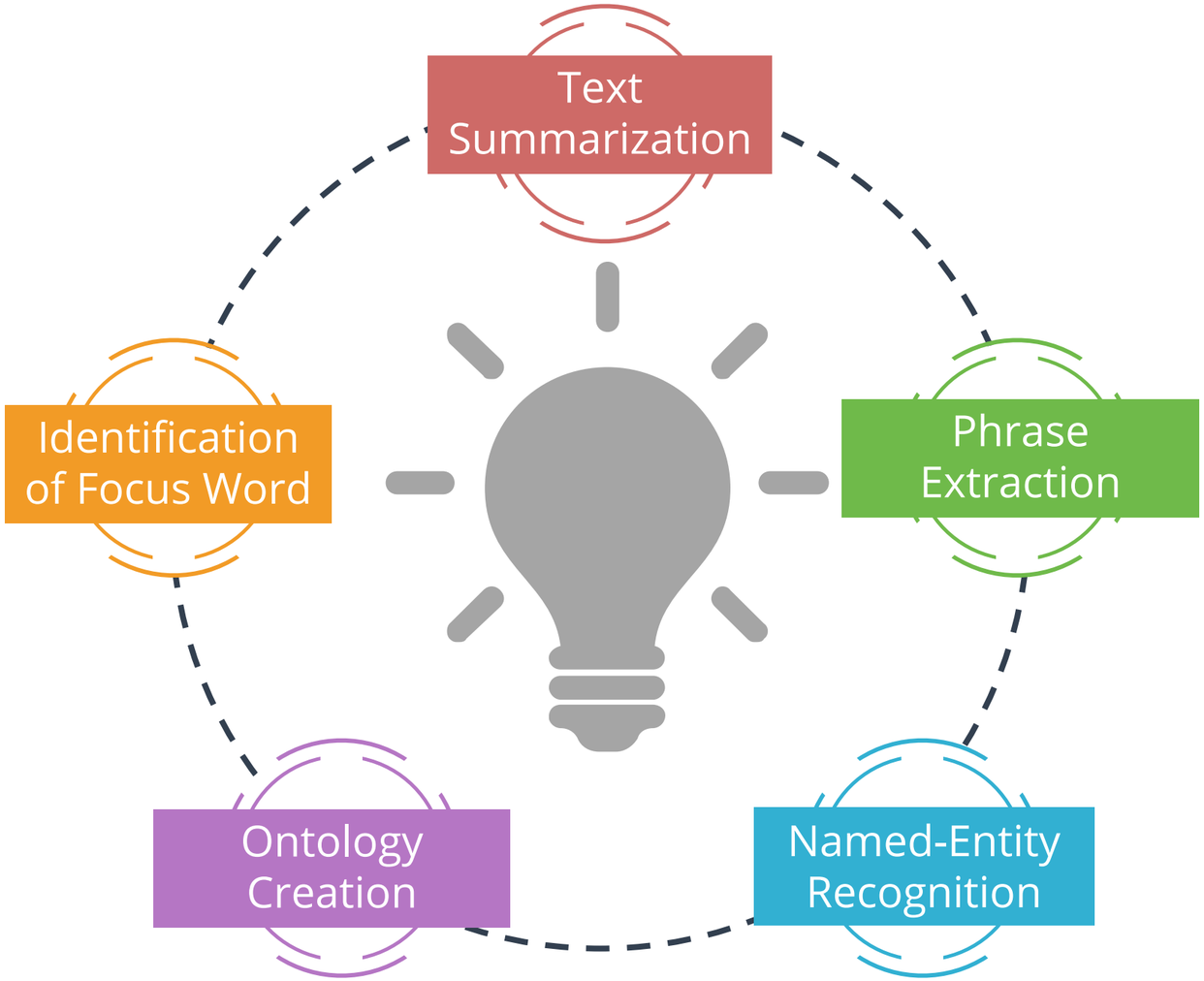
- The word “she” is personal pronoun and noun phrase.
- VP is verb phrase and S is sentence. VP has three words: enjoys, playing, tennis.



©Simplilearn. All rights reserved.

simplilearn

Constituency Parsing: Usage



©Simplilearn. All rights reserved.

simplilearn

DATA AND
ARTIFICIAL INTELLIGENCE

Morphological Techniques

©Simplilearn. All rights reserved.

Morphological Techniques

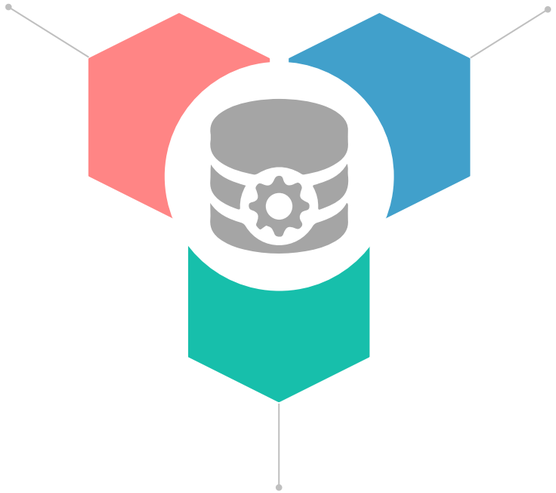
- Morphology is the study of structure and formation of words.
- Every important unit is called a morpheme.
- Example: UNHAPPINESS
  It has three parts and therefore, three morphemes: **UN HAPPY NESS**

©Simplilearn. All rights reserved.

Morphological Techniques: Classes

Classification of morphological structural types:

Isolating Languages



Agglutinative Languages

Inflecting Languages

©Simplilearn. All rights reserved.

simplilearn

Morphological Techniques: Process

©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE

Fuzzy Search

©Simplilearn. All rights reserved.

Fuzzy Search

1. Fuzzy search is a process that identifies the relevant web pages or texts for searched argument.
2. It is done by fuzzy match program which returns a list of results even though the search argument spelling is not correct.
3. Fuzzy string is also known as approximate string matching.

©Simplilearn. All rights reserved.

Fuzzy Search: Techniques

Regular Expression

Edit Distance

Wildcard in SQL Matching

©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE

Document and Sentence Similarity

©Simplilearn. All rights reserved.

Document and Sentence Similarity

It determines the similarity between the two pieces of text or documents and also, used to cluster the document.

The following are the types of similarity determined:

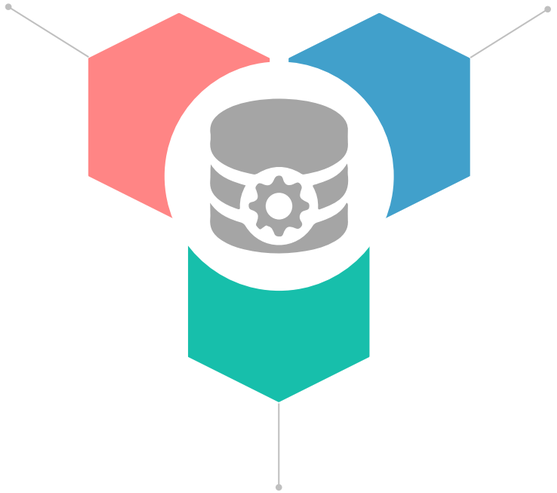
Lexical

Semantic

©Simplilearn. All rights reserved.

Document and Sentence Similarity: Techniques

LSI + Cosine Similarity



Jaccard Similarity

Word embeddings + Word Mover Distance

©Simplilearn. All rights reserved.

Document and Sentence Similarity: Example

- "The mouse ate the cat's food"
- "The cat ate the mouse"

How similar are these?

- On Lexical similarity, they are very close.
- On Semantic similarity, they convey totally different meanings.

©Simplilearn. All rights reserved.

Document and Sentence Similarity: Applications

Quora

stackoverflow

Customer Service

©Simplilearn. All rights reserved.

Document Indexing

©Simplilearn. All rights reserved.

Document Indexing

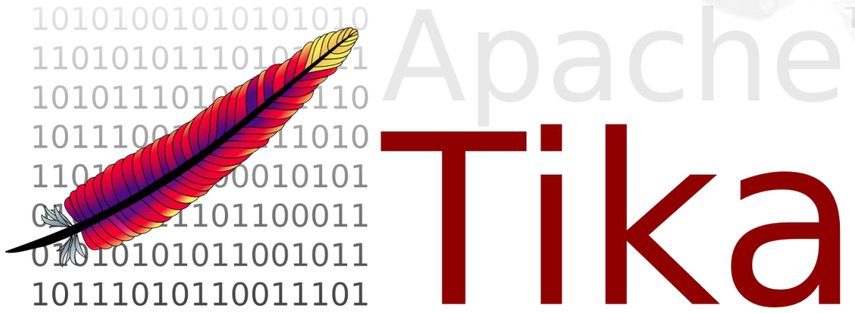
1. It is the process of associating or tagging document or file with various search terms.
2. It is used for search and retrieval purpose in the future.
3. It creates searchable information.
4. Index information is stored in documents or records management system.
5. **Full Text Indexing** and **Date-Based Indexing** are the two types of indexing.

©Simplilearn. All rights reserved.

simplilearn

Document Indexing: Tools

elasticsearch



©Simplilearn. All rights reserved.

Document Indexing: Use Cases

Searching books in digital library

Searching the relevant web page on search engine

Document tracking system

©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE

Sentiment Analysis

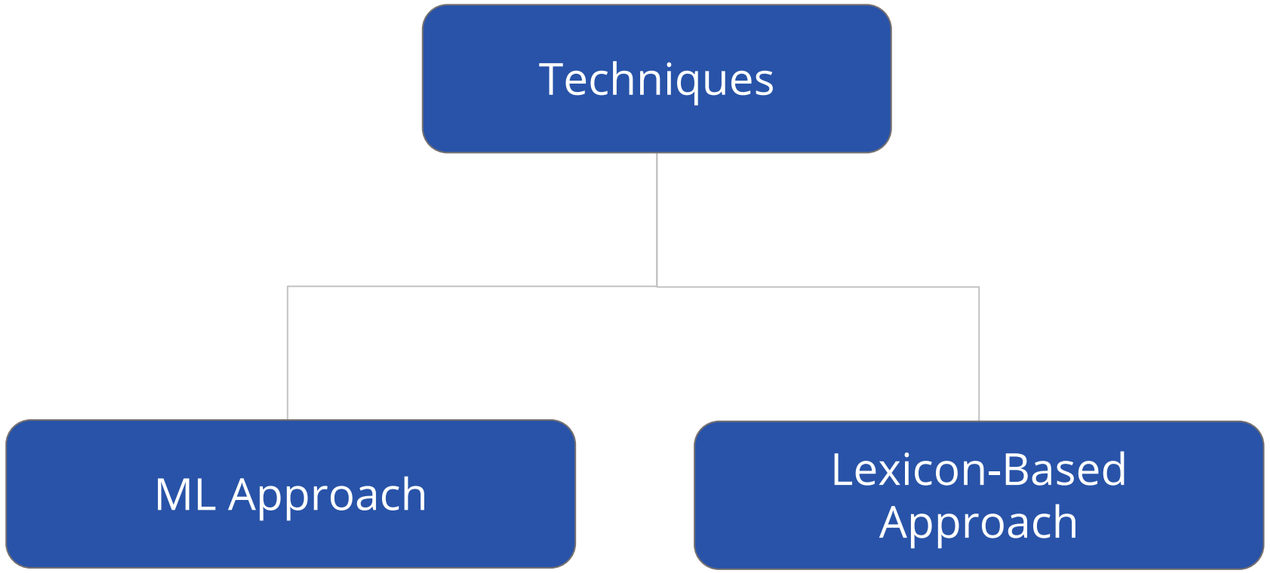
©Simplilearn. All rights reserved.

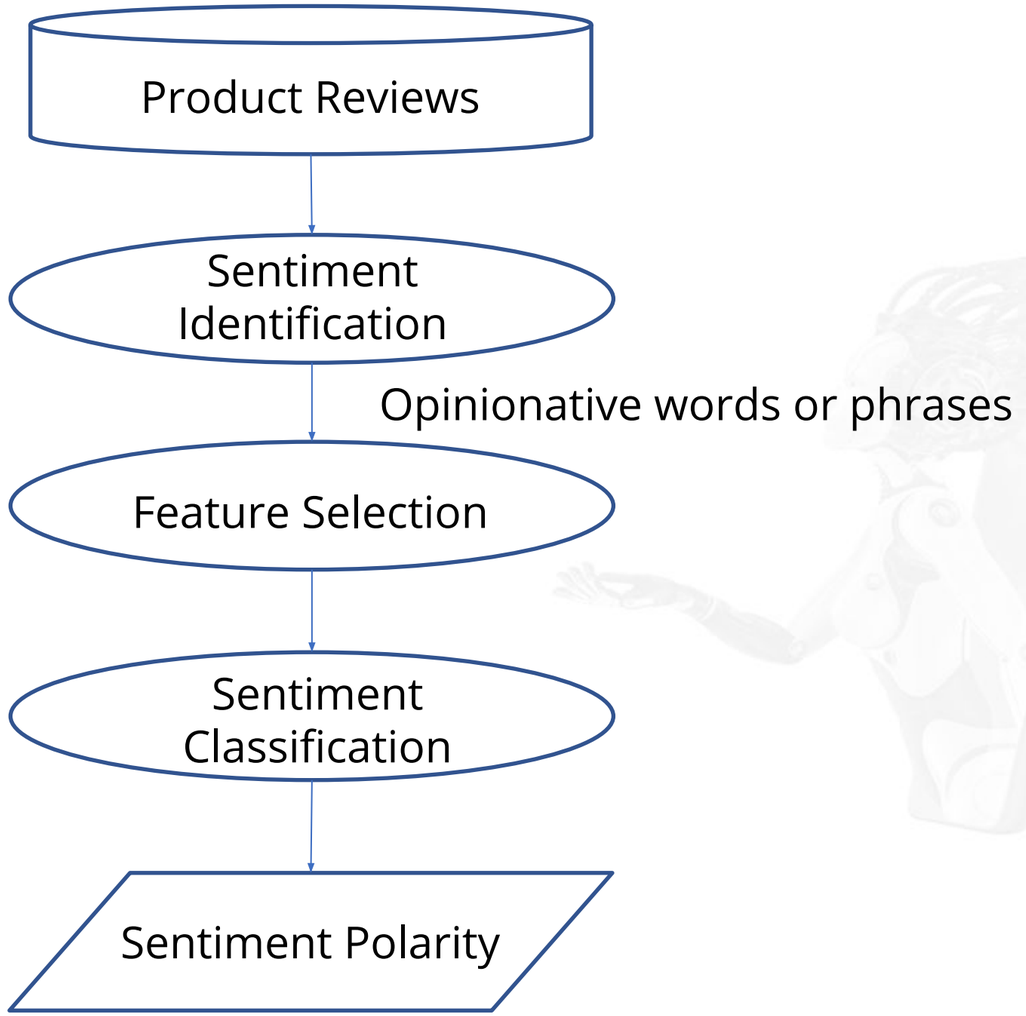
Sentiment Analysis

- Sentiment analysis is a technique used for text analysis using NLP or ML to assign weighted sentiment scores to the opinion mentioned in the text.
- Sentiment is a view or opinion that is held or expressed.
- It is widely applied to review and survey responses online and on social media, for providing good customer service and marketing.

©Simplilearn. All rights reserved.

Sentiment Analysis: Techniques





©Simplilearn. All rights reserved.

simplilearn

Sentiment Analysis: Techniques

Below are the steps involved in sentiment analysis:

1. User gives the feedback or comment on the portal
2. Categorization of possible sentiments is done
3. Feature selection is applied after noise removal and text-processing
4. ML or Lexicon-based approach will be applied for sentiment classification
5. Score polarity is defined

©Simplilearn. All rights reserved.

simplilearn

Sentiment Analysis: Applications

©Simplilearn. All rights reserved.

simplilearn

Get the Polarity of the Given Survey for a Product

Problem Statement: Sentiment analysis is one of the most common applications in natural language processing. With sentiment analysis, we can decide with what emotion a text is written. Using sentiment analysis, get the polarity of the given survey for a product.

Access: Click on the **Practice Labs** tab on the left side panel of the LMS. Copy or note the username and password that is generated. Click on the **Launch Lab** button. On the page that appears, enter the username and password in the respective fields, and click **Login**.

ASSISTED PRACTICE

©Simplilearn. All rights reserved.

Extract City and Person Name from Text

Problem Statement: While conversing with other users, we try to understand what is being said and the entities mentioned in the sentence. Machine also has the capability to understand these entities and we can achieve this by using named-entity recognition. Apply existing NER library to extract city and person name from the given text.

Access: Click on the **Practice Labs** tab on the left side panel of the LMS. Copy or note the username and password that is generated. Click on the **Launch Lab** button. On the page that appears, enter the username and password in the respective fields, and click **Login**.

ASSISTED PRACTICE

©Simplilearn. All rights reserved.

Identifying Top Product Features from User Reviews

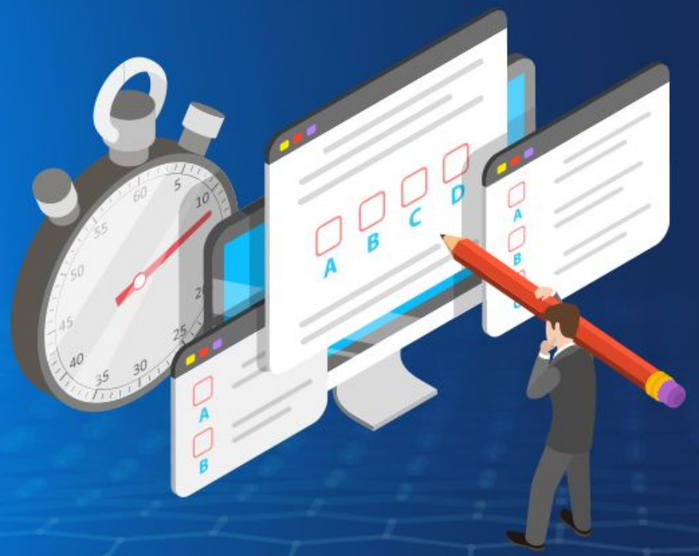
Objective: Analyze the reviews of a product on Amazon to identify the features of the product that customers are talking about the most. Use POS tagging and extracting the right parts of speech to analyze.

Problem Statement: Amazon Tap is a product made by Amazon. It is a portable Bluetooth and Wi-Fi-enabled speaker that gives you rich, full-range sound. Customers can use Alexa Voice Service that performs the role of a virtual assistant. Amazon is planning to launch a new version of Tap but, need to understand what features of the products are important for the customers. Amazon wants to analyze the reviews gathered on the product so far, to assess which of the product's features are most talked about by the customers.

LESSON-END PROJECT

©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE



Knowledge Check

©Simplilearn. All rights reserved.

Knowledge Check

1

Dependency parsing is used for _____.

a. Semantic understanding

b. Finding relation between words

c. Understanding modifiers

d. All of the above

©Simplilearn. All rights reserved.

Knowledge Check

1

Dependency parsing is used for _____.

a. Semantic understanding

b. Finding relation between words

c. Understanding modifiers

d. All of the above

The correct answer is **d.**

It helps in finding semantic structure with relation and modifiers.

©Simplilearn. All rights reserved.

Knowledge Check

2

I would like to extract the person's name from the paragraph. Which of the following techniques is suitable?

a. POS Tagging

b. Regular Expression

c. Word Sense Disambiguation

d. Named-Entity Recognition

©Simplilearn. All rights reserved.

Knowledge Check

2

**I would like to extract the person's name from the paragraph.
Which of the following techniques is suitable?**

a. POS Tagging

b. Regular Expression

c. Word Sense Disambiguation

d. Named-Entity Recognition

The correct answer is **d.**

Named-entity recognition understands the names and gives you the named-entity words.

©Simplilearn. All rights reserved.

Knowledge Check

3

POS tags are the parts of speech tags which indicate the tag only for _____.

a. Single word

b. Consecutive word

c. N-gram phrases

d. Complete sentence

©Simplilearn. All rights reserved.

Knowledge Check

3

POS tags are the parts of speech tags which indicate the tag only for _____.

a. Single word

b. Consecutive word

c. N-gram phrases

d. Complete sentence

The correct answer is **a.**

POS tag indicates the tag only for single words. The combined tags are only for phrases.

©Simplilearn. All rights reserved.

Knowledge Check

4

“Sky is Blue". Which technique is best suitable to understand the meaning of "Blue"?

a. Word Sense Disambiguation

b. Dependency Parsing

c. Fuzzy Search

d. Coreference Resolution

©Simplilearn. All rights reserved.

simplilearn

Knowledge Check

4

“Sky is Blue". Which technique is best suitable to understand the meaning of "Blue"?

a. Word Sense Disambiguation

b. Dependency Parsing

c. Fuzzy Search

d. Coreference Resolution

The correct answer is **a.**

A word has many meanings in different contexts. WSD has the technique to overcome this problem.

©Simplilearn. All rights reserved.

Knowledge Check

5

What is the main application of sentiment analysis?

a. Understanding user feedback

b. Understanding syntactic structure

c. Fuzzy Search

d. Coreference Resolution

©Simplilearn. All rights reserved.

Knowledge Check

5

What is the main application of sentiment analysis?

a. Understanding user feedback

b. Understanding syntactic structure

c. Fuzzy Search

d. Coreference Resolution

The correct answer is **a.**

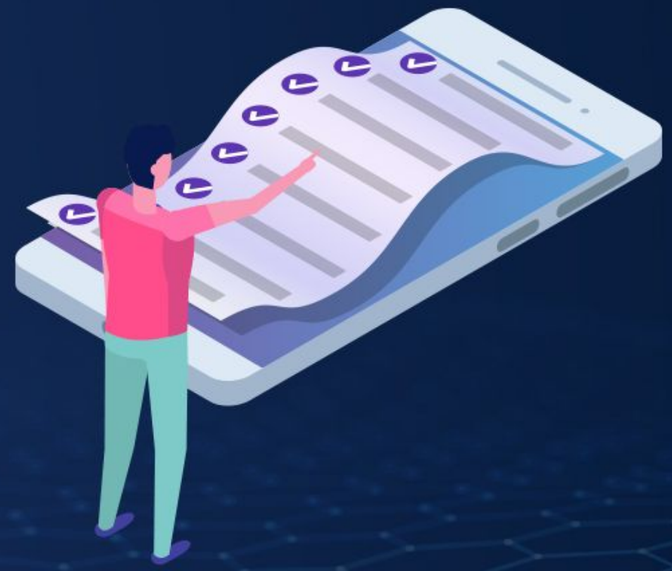
User's feedback is used to improve services and sentiment analysis plays a major part in this process.

©Simplilearn. All rights reserved.

Key Takeaways

You are now able to:

- Define parts-of-speech tagging
- Explain the different parsing methods
- Apply fuzzy search to identify similar words
- Get the polarity of the given survey for a product
- Extract city and person name from a text



©Simplilearn. All rights reserved.