AWS Academy Machine Learning Foundations

Module 6: Introducing Natural Language Processing

Module overview

Sections

1. Overview of natural language processing (NLP)
2. Natural language processing managed services
3. Module wrap-up

Demonstrations

- Introducing Amazon Polly
- Introducing Amazon Comprehend
- Introducing Amazon Translate

Lab

- Guided Lab: Creating a Bot to Schedule Appointments

Knowledge check

Module objectives

At the end of this module, you should be able to:

- Describe the natural language processing (NLP) use cases that are solved by using managed Amazon ML services
- Describe the managed Amazon ML services available for NLP
- Use managed Amazon ML Services

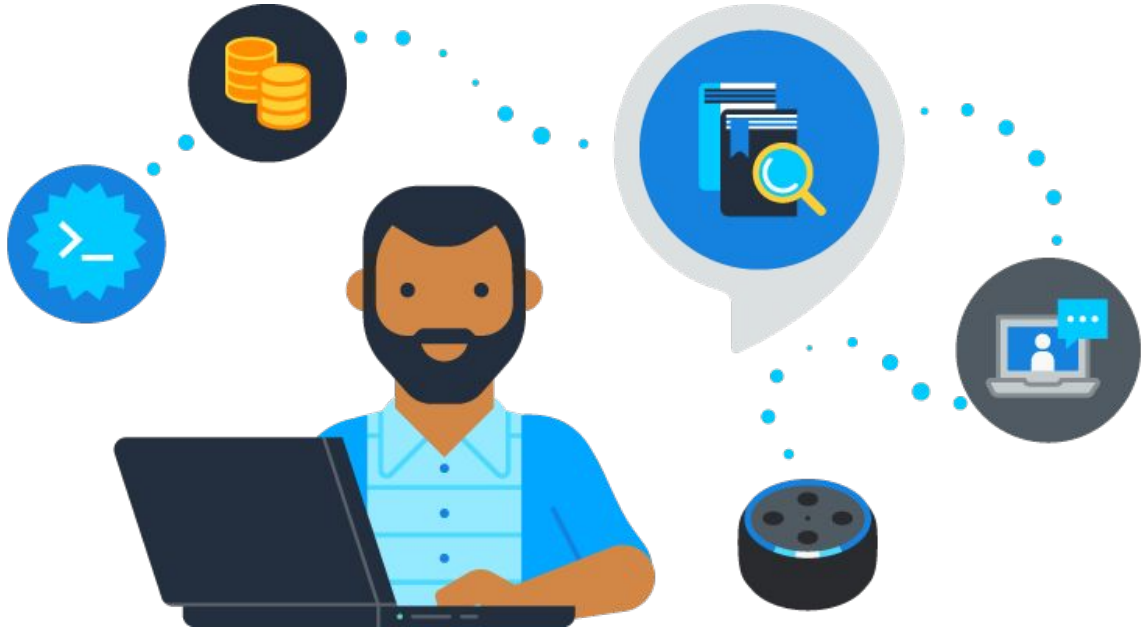
Module 6: Introducing Natural Language Processing

Section 1: Overview of natural language processing

aws academy

Natural language processing (NLP)

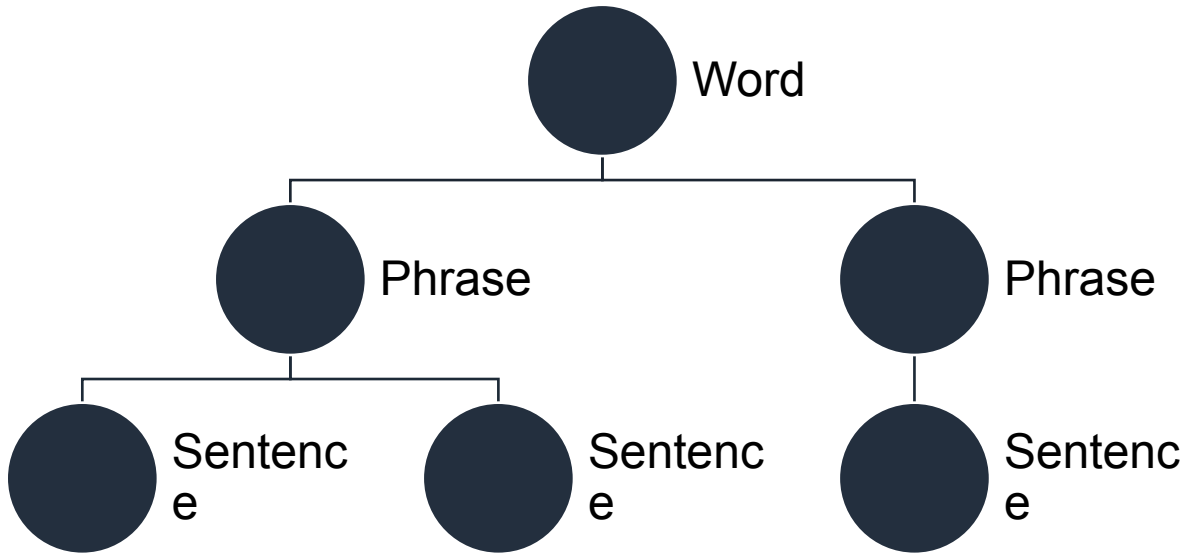
"Alexa, what's it like outside?"



What is NLP?

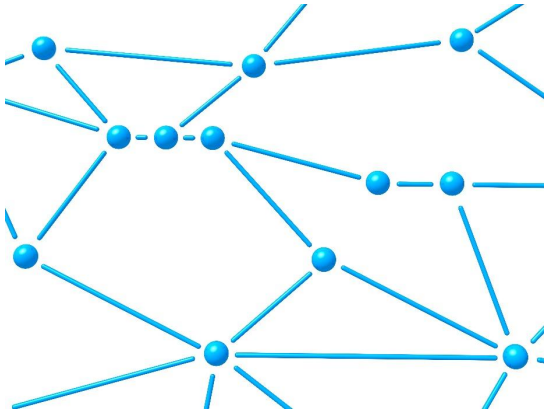
NLP develops computational algorithms to automatically analyze and represent human language.

By evaluating the structure of language, machine learning systems can process large sets of words, phrases, and sentences.



NLP challenges

Lack of precision



Many complex dependencies

Meaning that is based on context

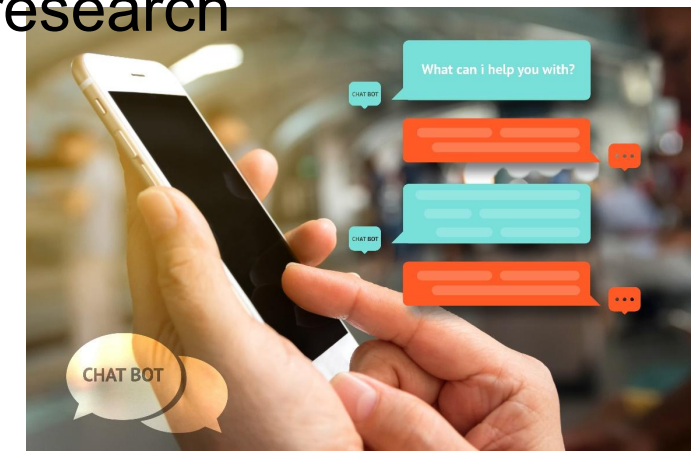
Lack of structure

Natural language processing use cases

Search applications

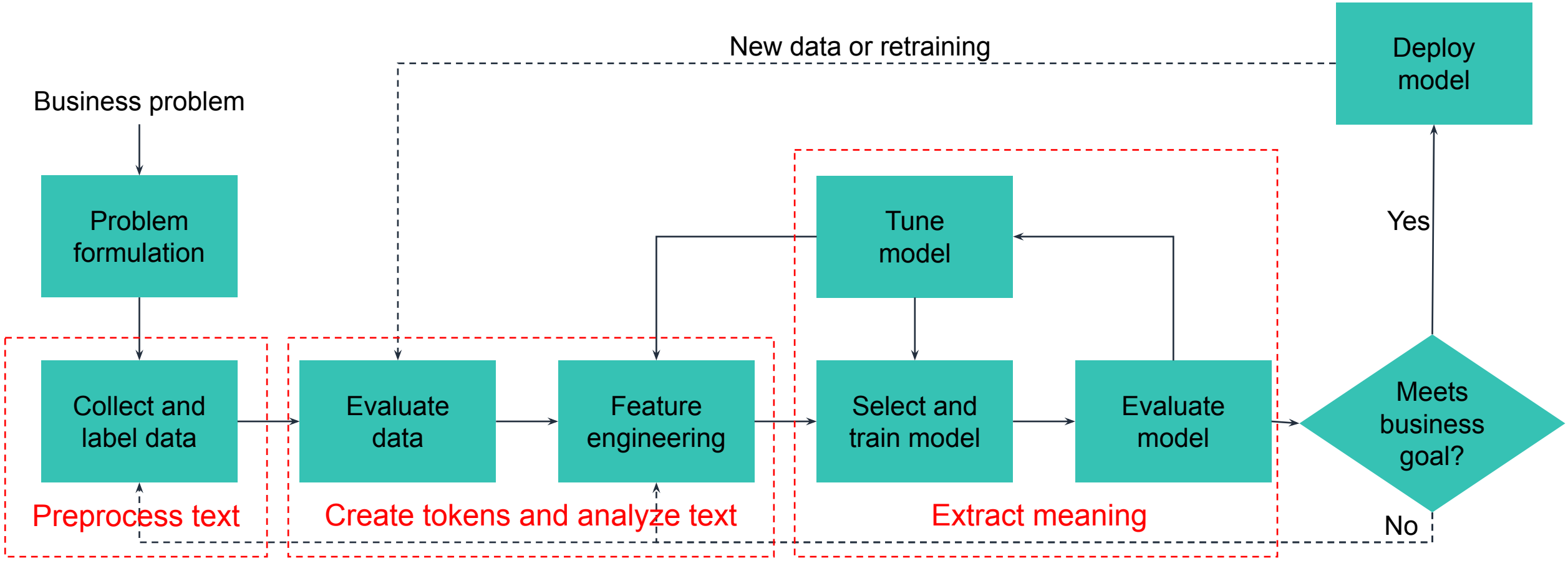
Market and social research

Human machine interfaces



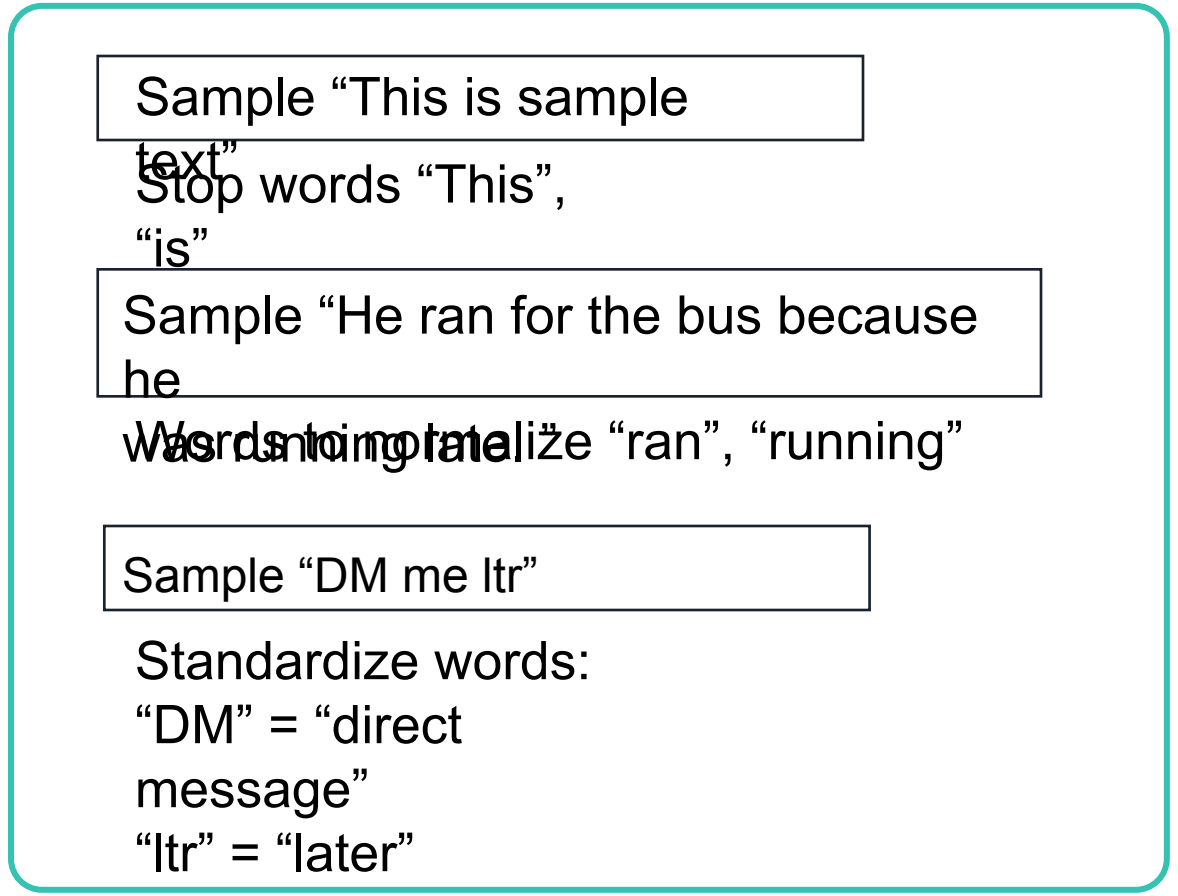
Chatbots

Natural language processing flow



Preprocessing text

- Common preprocessing steps –
  - Remove stop words
  - Normalize similar text
  - Standardize unrecognized text
- Other preprocessing steps –
  - Encoding
  - Spelling and grammar checks
- Multiple libraries and tools are available for preprocessing (for example, NLTK for Python)



Sample Preprocessing

Creating tokens and feature engineering

- Load data by using tokens
  - You can use tokens to convert words into items in a DataFrame
- Develop features by applying a model
  - Common models include *bag of words* and *term frequency and inverse document frequency (TF-IDF)*

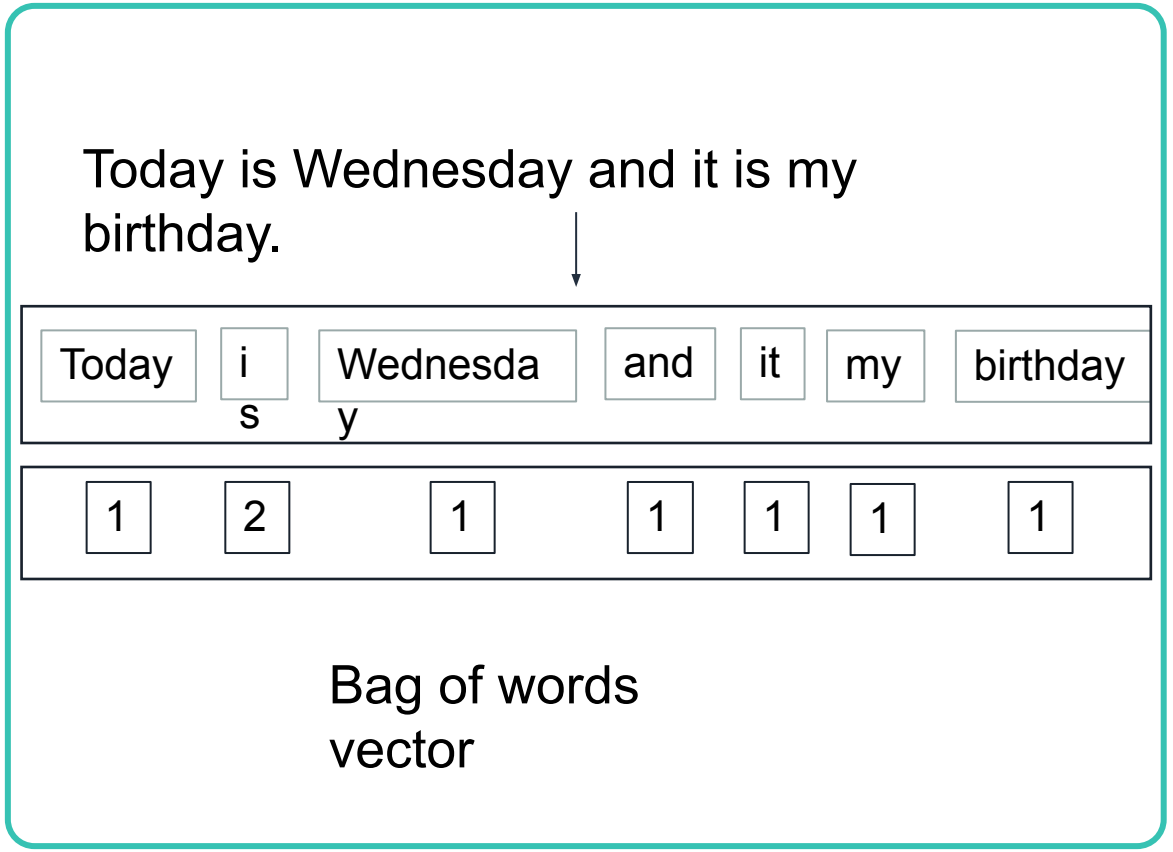
```
from nltk.tokenize import word_tokenize
text = “this is some sample text.”
Print(word_tokenize(text))

Output: [‘this’,‘is’,‘some’,‘sample’,‘text’ ‘.’]
```

Sample token code

Example NLP model: Bag of words

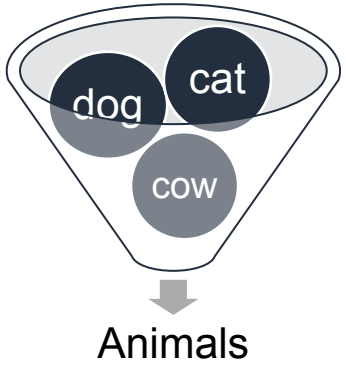
- Create a vector for each sentence or phrase
- Evaluate words in a sentence that is based on frequency
  - Frequency creates a vector for each sentence or phrase



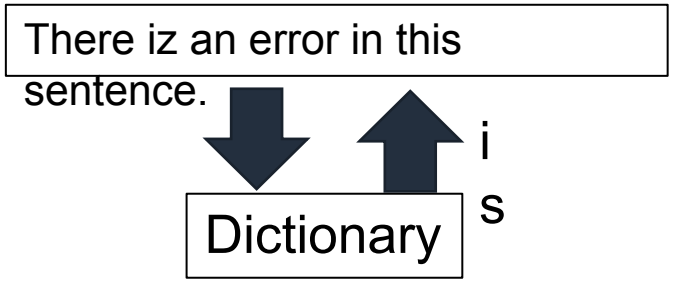
Example NLP Model

Text analysis categories

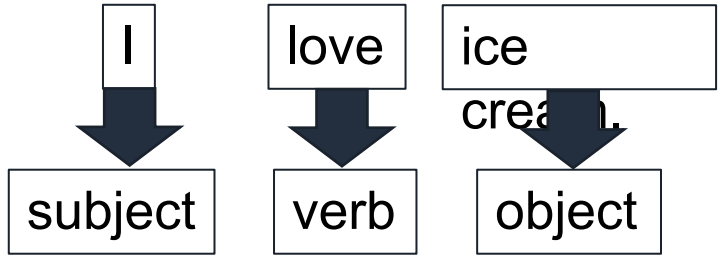
Classifying text



Discovering similarities



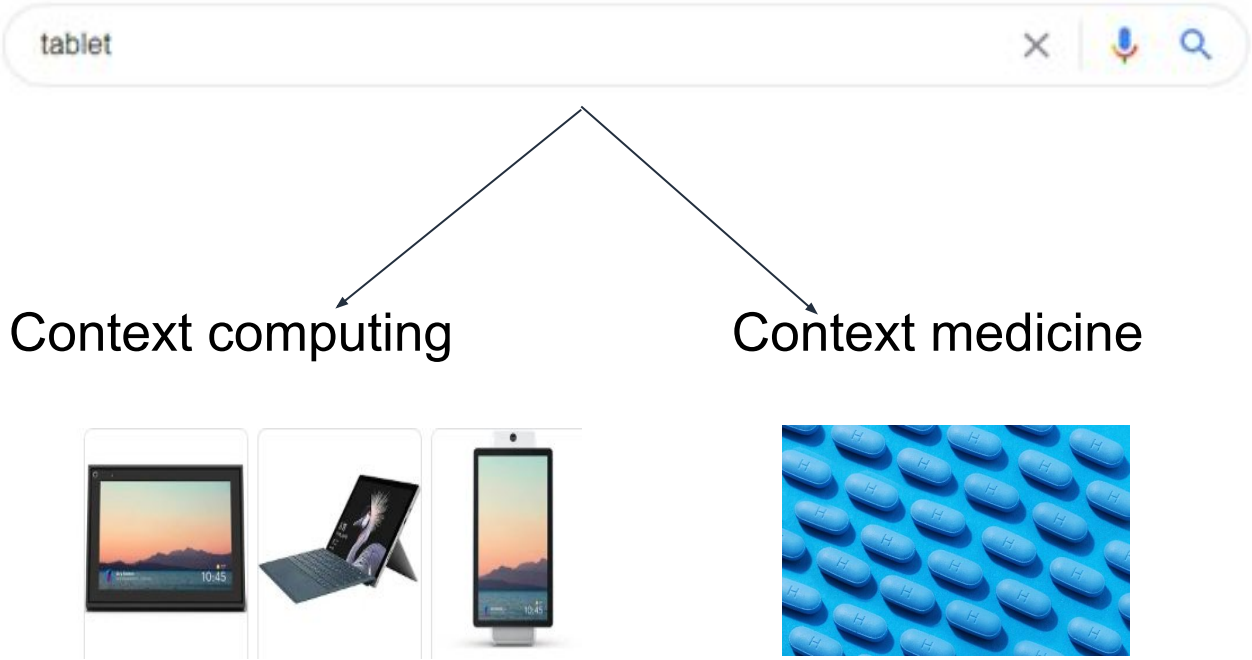
Deriving relationships



Capture context

Understanding context for the text is a major challenge for NLP:

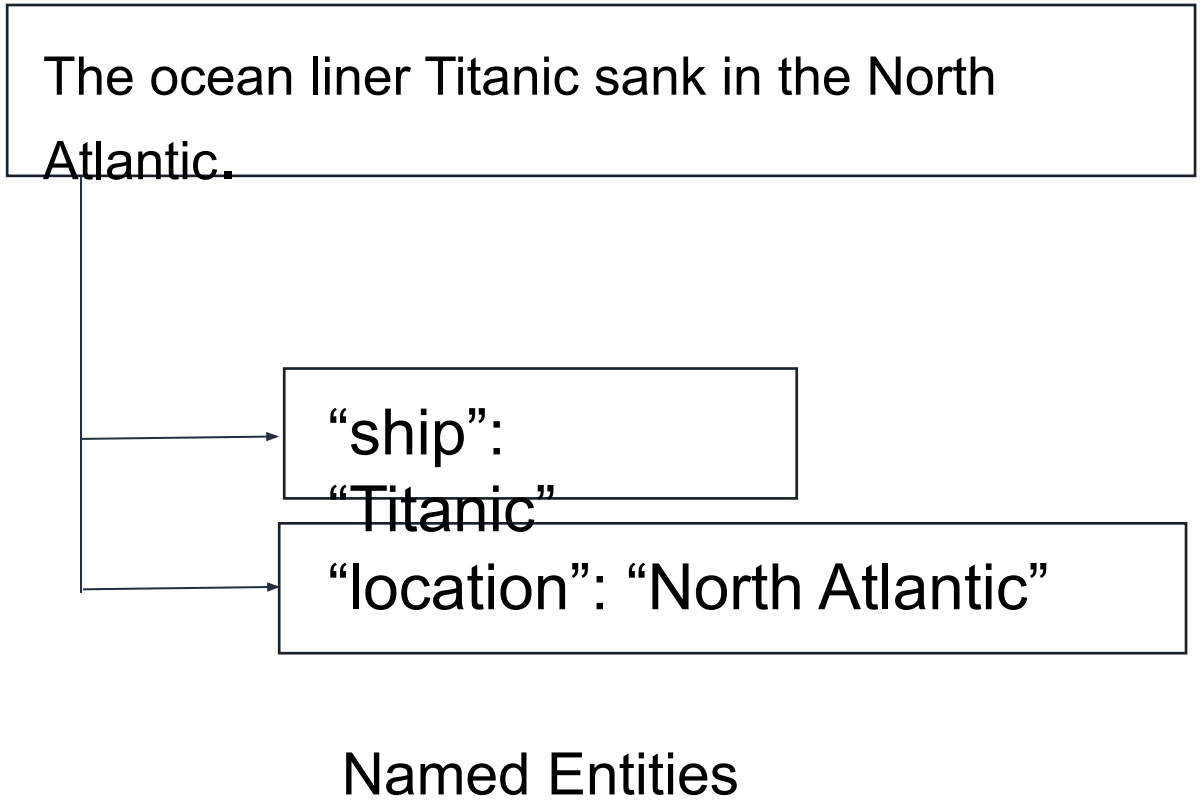
- Tagging words with the appropriate part of speech helps to capture context
- NLP libraries provide token functions to help with tagging



Derive meaning by entity extraction

Extract entities by using named entity recognition (NER):

- Identify noun phrases
- Classify phrases by using a classification algorithm
- Disambiguate entities by using a knowledge graph



Section 1 key takeaways

- As a domain, NLP predates machine learning
- NLP development maps directly to the ML development process
- Some of the main use cases for NLP are search query analysis, human-machine interaction, and market or social research
- NLP is difficult because of the imprecise nature of human language

Module 6: Introducing Natural Language Processing

Section 2: Natural language processing managed services

aws academy

Amazon Transcribe

Amazon Transcribe

Amazon Transcribe is a fully managed service that uses advanced machine learning technologies to recognize speech in audio files and transcribe them into text. You can use Amazon Transcribe to convert audio to text and to create applications that incorporate the content of audio files.

- Recognize recorded voices
- Convert streaming audio to text
- Customize specialized vocabularies
- Integrate with applications by using WebSockets
- Build subtitles for multiple languages in real time

Amazon Transcribe use cases

Medical transcription

Subtitles

Streaming content labeling

Call center monitoring

Amazon Polly

Amazon Polly is a managed service that converts text into lifelike speech. Amazon Polly supports multiple languages and includes various lifelike voices.

- Generate voice from plain text or Speech Synthesis Markup Language (SSML) format
- Create output in multiple audio formats
- Offers a pay-for-use policy and uses AWS infrastructure to keep costs low

Amazon Polly use cases

News service production

Language training

Navigation systems

Animation production

Demonstration: Introducing Amazon Polly

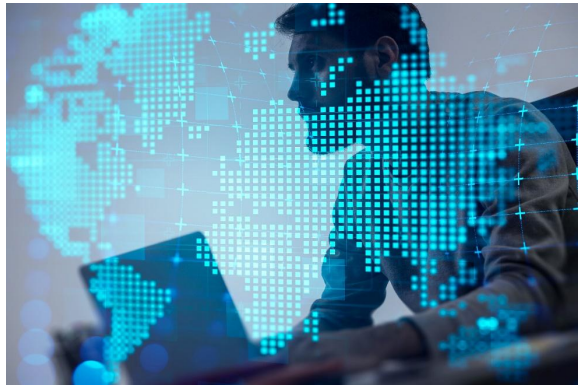
Amazon Translate

Amazon Translate

Amazon Translate is a fully managed text translation service that uses advanced machine learning technologies to provide high-quality translation on demand.

- Develop multilingual user experiences for your applications
- Translate documents to multiple languages
- Analyze incoming text in multiple languages

Amazon Translate use cases



International websites

Software localization

Multilingual chatbots

International media

Demonstration: Introducing Amazon Translate

Amazon Comprehend

Amazon Comprehend

Amazon Comprehend uses NLP to extract insights about the content of documents. It develops insights by recognizing the entities, key phrases, language, sentiments, and other common elements in a document.

- Extract key entities from a document, such as people or locations
- Identify the language that is used in a document
- Determine the sentiment—such as positive, negative, neutral, or mixed—that is expressed in a document
- Identify the part of speech for individual words in a document

Amazon Comprehend use cases

Document analysis

Fraud detection

Mobile app analysis

Content management

Demonstration: Introducing Amazon Comprehend

Amazon Lex

Amazon Lex

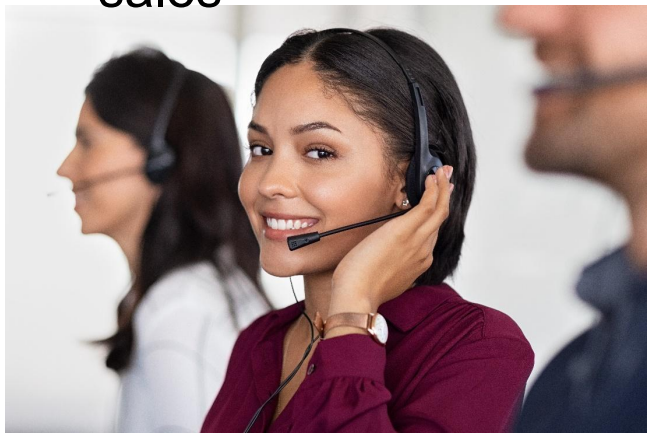
Amazon Lex is an AWS service for building conversational interfaces for applications by using voice and text. With Amazon Lex, the same conversational engine that powers Amazon Alexa is now available to any developer.

- Build a chatbot that can interact with voice and text to ask questions, get answers, or complete tasks
- Automatically scale your chatbot with AWS Lambda
- Store log files of conversations for analysis

Amazon Lex use cases

Inventory and sales

Interactive assistants



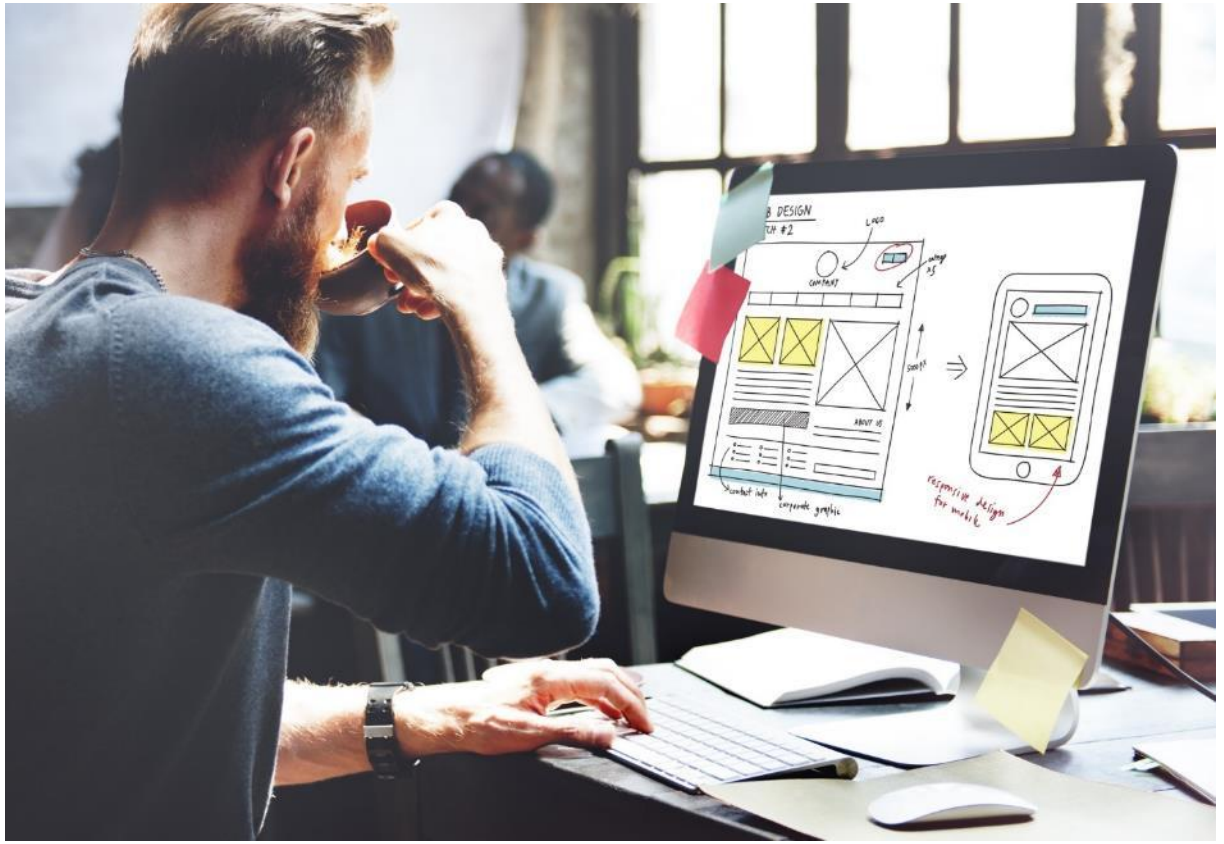
Customer service interfaces

Database queries

Section 2 key takeaways

- Amazon Transcribe can automatically convert spoken language to text
- Amazon Polly can convert written text to spoken language
- Amazon Translate can create real-time translation between languages
- Amazon Comprehend automates many of the NLP use cases that are reviewed in this module
- Amazon Lex can create a human-like interface to your applications

Module 6 – Guided Lab: Creating a Bot to Schedule Appointments



Module 6: Introducing Natural Language Processing

Module wrap-up

aws academy

Module summary

In summary, in this module you learned how to:

- Describe the natural language processing (NLP) use cases that are solved by using managed Amazon ML services
- Describe the managed Amazon ML services available for NLP
- Use managed Amazon ML Services

Complete the knowledge check

Additional resources

- What is Amazon Comprehend?
- What is Amazon Polly?
- What is Amazon Lex?
- What is Amazon Transcribe?
- What is Amazon Translate?

Thank you

aws academy