

APLICACIONES DEL PROCESAMIENTO NATURAL DEL LENGUAJE (NLP)

DR. MARTÍN BILBAO

DIRECTOR DEL GRUPO DE INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL

DEPARTAMENTO DE INFORMÁTICA - UNPSJB

¿QUÉ ES PROCESAMIENTO NATURAL DEL LENGUAJE?

- DESARROLLAR APLICACIONES Y SERVICIOS QUE PUEDAN RECONOCER EL LENGUAJE HUMANO
- DONDE SE APLICA:

Reconocimiento
de voz

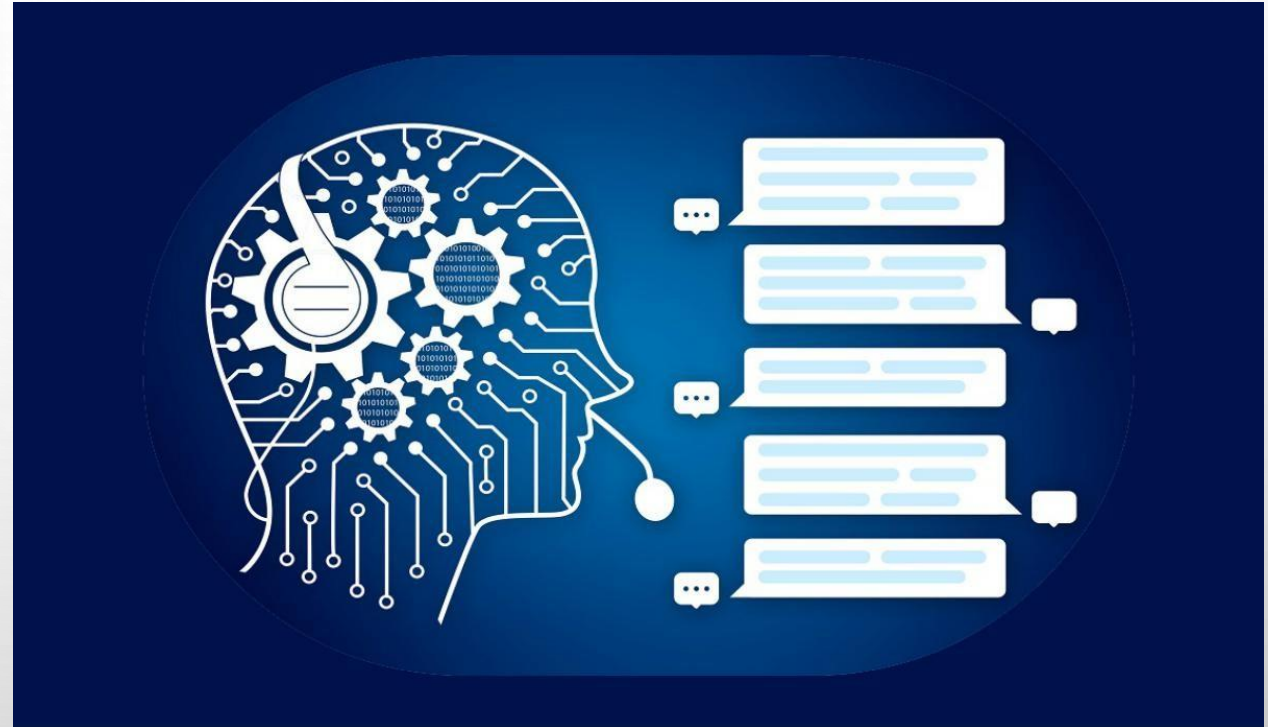
Análisis de
sentimientos

NLP

Traducción
del habla

Chatbots

APLICACIONES DEL PNL



APLICACIONES DEL PROCESAMIENTO NATURAL DEL LENGUAJE

- Búsqueda web
- Búsqueda de palabras claves
- Filtros de SPAM
- Visión de Información
- Filtrado de información

Clasificación
de textos

Análisis de
sentimientos

- Valoración de productos
- Valoración de mensajes de negocio
- Análisis de popularidad

NLP

- Transcribir el habla
- Órdenes por voz a sistemas electrónicos
- Chatbots potenciados

Reconocimiento
y síntesis de
voz

Chatbots

- Atención al cliente virtual
- Respuestas automáticas



PARTE I: SISTEMA DE CLASIFICACIÓN DE TEXTO

DR. MARTÍN BILBAO

DIRECTOR DEL GRUPO DE INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL

DEPARTAMENTO DE INFORMÁTICA - UNPSJB

¿QUÉ ES LA CLASIFICACIÓN DE TEXTOS?

- CATEGORIZAR Y ORGANIZAR UN TEXTO PARA AYUDAR EN LA BÚSQUEDA, FILTRADO Y ANÁLISIS DE INFORMACIÓN



Natural Language Toolkit (NLTK): Es la biblioteca más popular y sencilla para el procesamiento del lenguaje natural (NLP) operativa para Python con una gran comunidad.



APLICACIONES DEL PNL

PASOS EN LA CLASIFICACIÓN DE TEXTO

- PASO 1: TOKENIZAR TEXTO

1. Tokenizar
Texto



Dividir el texto en tokens (palabras o sentencias)

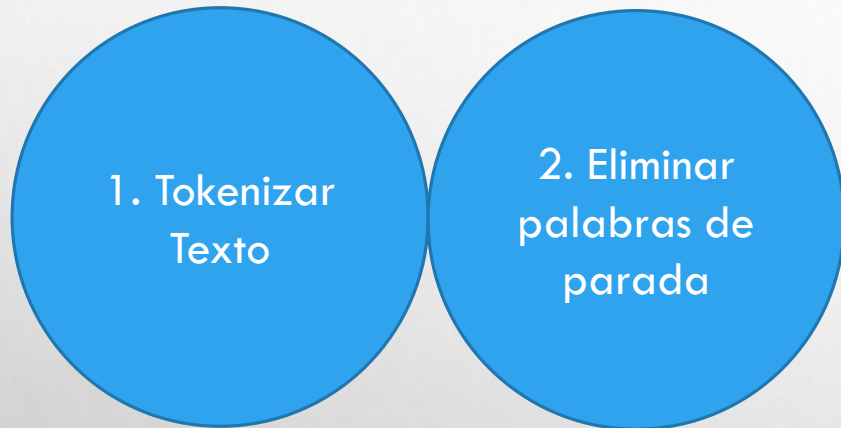
Juan es estudiante de Informática



Juan – es – estudiante – de – Informática (5 tokens)

PASOS EN LA CLASIFICACIÓN DE TEXTO

- PASO 2: ELIMINAR PALABRAS DE PARADA



Eliminar palabras poco relevantes (artículos, proposiciones etc.)

Juan es estudiante de Informática



Juan – es – estudiante – Informática (4 tokens)

PASOS EN LA CLASIFICACIÓN DE TEXTO

- PASO 3: OBTENER SINÓNIMOS Y ANTÓNIMOS

1. Tokenizar
Texto

2. Eliminar
palabras de
parada

3. Obtener
sinónimos y
antónimos



Wordnet es una base de datos dentro de NLTK

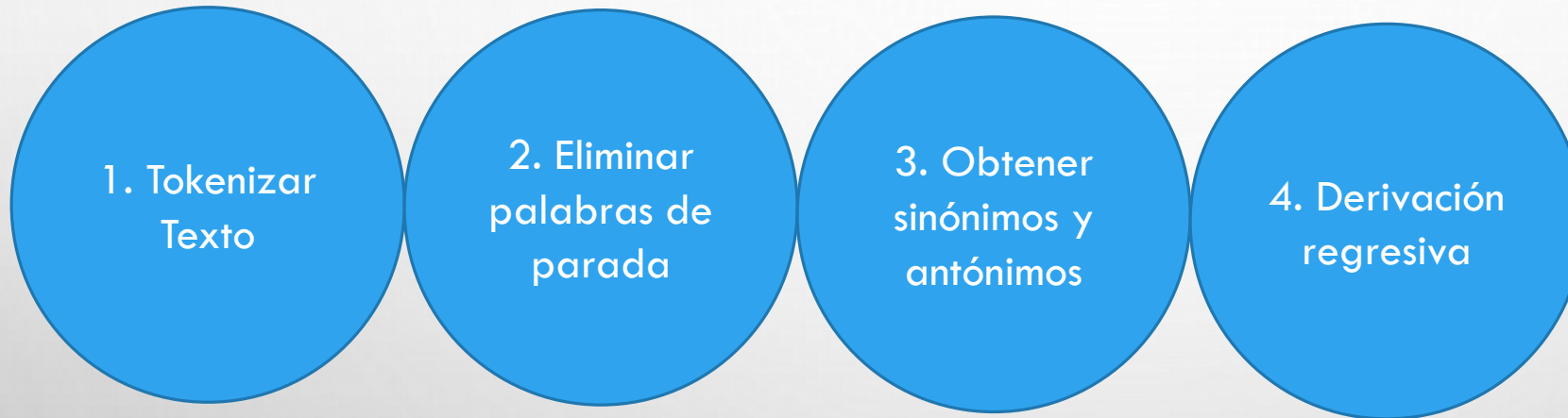
Se obtienen sinónimos para computar conjuntamente la frecuencia de aparición y su clasificación

Se obtienen antónimos para evaluar el contrapeso de cada palabra (útil en el análisis de sentimiento, por ejemplo éxito<->fracaso).

APLICACIONES DEL PNL

PASOS EN LA CLASIFICACIÓN DE TEXTO

- PASO 4A: DERIVACIÓN REGRESIVA



La derivación regresiva se utiliza para encontrar la palabra raíz eliminando afijos (prefijos o sufijos)

Optimiza la agrupación y clasificación de textos

Algoritmo Porter de derivación regresiva (PorterStemmer) por defecto en inglés en NLTK

SnowballStemmer es la versión en español en NLTK

APLICACIONES DEL PNL

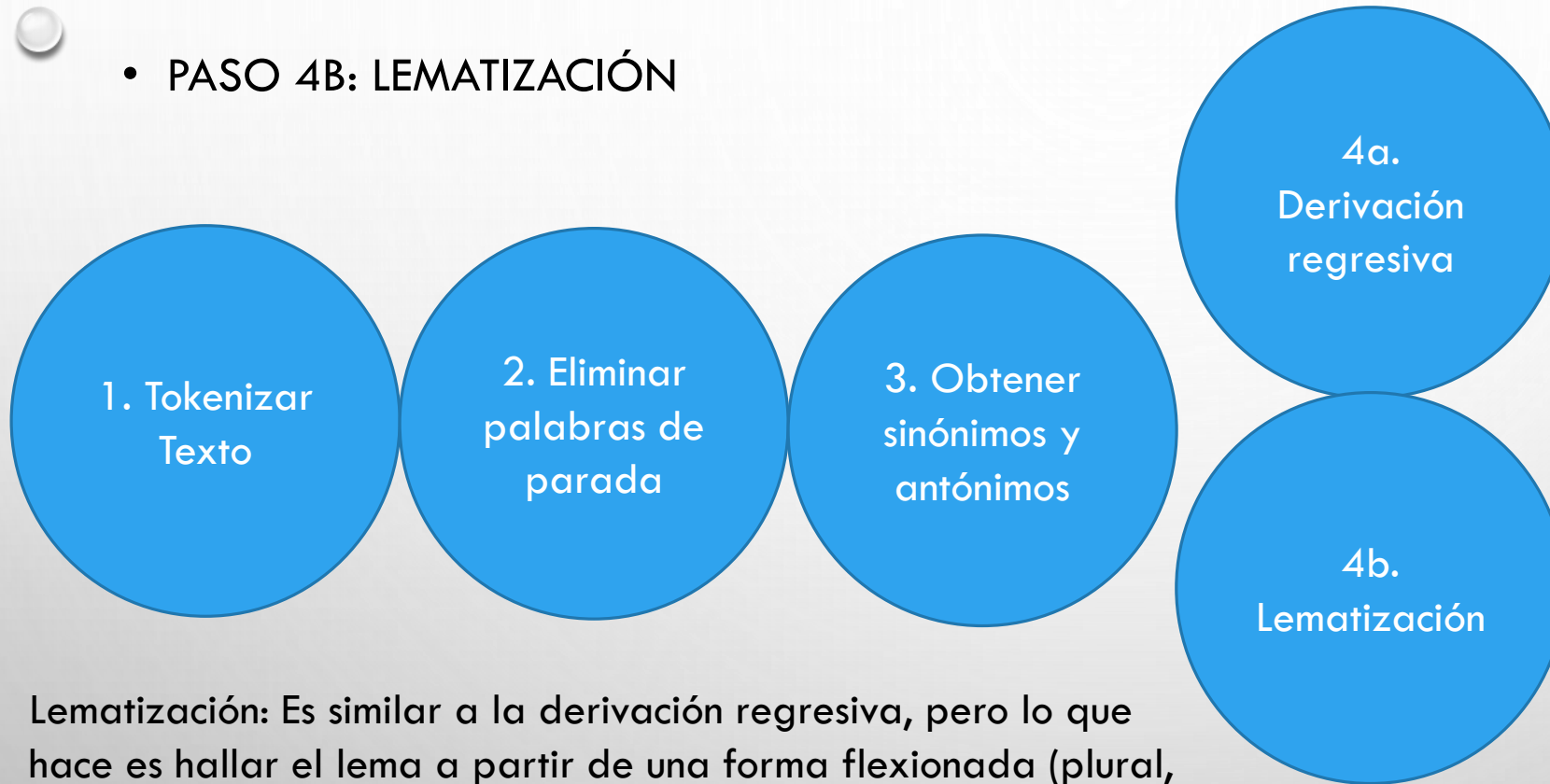
Participando
Participante



Participan

PASOS EN LA CLASIFICACIÓN DE TEXTO

- PASO 4B: LEMATIZACIÓN



Lematización: Es similar a la derivación regresiva, pero lo que hace es hallar el lema a partir de una forma flexionada (plural, verbo conjugado, femenino, masculino).

APLICACIONES DEL PNL

Dije
Mesas



Decir
Mesa

DIFERENCIA ENTRE LEMATIZACIÓN Y DERIVACIÓN REGRESIVA

- LA DERIVACIÓN REGRESIVA FUNCIONA CON PALABRAS SIN CONOCER SU CONTEXTO Y LA RAÍZ OBTENIDA TIENE UNA PRECISIÓN MENOR PERO ES MÁS RÁPIDA QUE LA LEMATIZACIÓN.
- LA LEMATIZACIÓN OBTIENE COMO LEMA UNA PALABRA DEL MUNDO REAL.
 - SI SE NECESITA VELOCIDAD → DERIVACIÓN REGRESIVA
 - SI SE NECESITA PRECISIÓN → LEMATIZACIÓN



PARTE II: ANÁLISIS DE SENTIMIENTOS

DR. MARTÍN BILBAO

DIRECTOR DEL GRUPO DE INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL

DEPARTAMENTO DE INFORMÁTICA - UNPSJB

¿EN QUE CONSISTE EL ANÁLISIS DE SENTIMIENTOS

- COMPRENDER Y ANALIZAR LAS RESPUESTAS DE LAS PERSONAS SOBRE UN PRODUCTO, SERVICIO O ENTIDAD

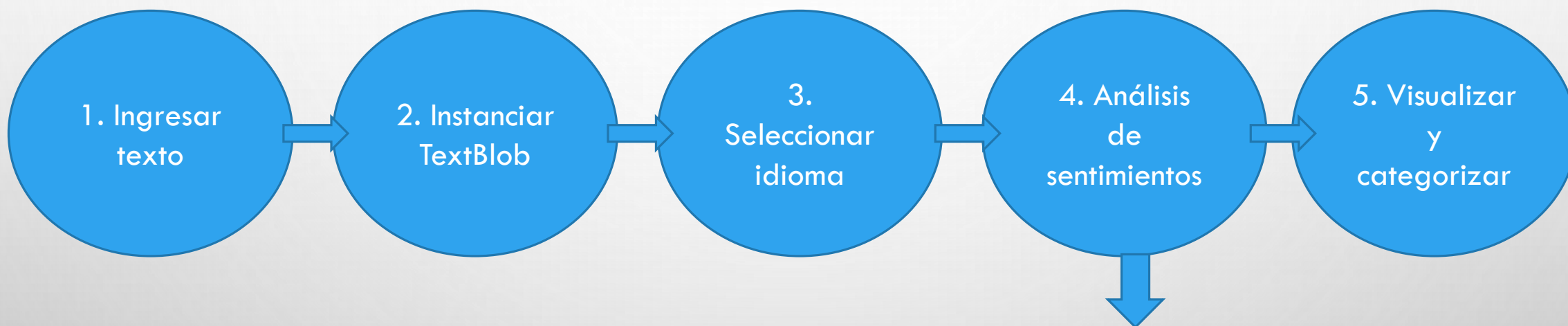
- OPINIONES
- EMOCIONES
- SENTIMIENTOS




- ESTA BASADO EN EL NLP Y LA ESTADÍSTICA ASIGNANDO VALORES:
 - TEXTO (POSITIVO, NEGATIVO, NEUTRAL), ESTO SE CONOCE COMO POLARIDAD (VALOR ENTRE -1 Y 1)
 - IDENTIFICAR EL SENTIMIENTO (CONTENTO, TRISTE, ENFADADO ETC.)
 - ANÁLISIS DE LA SUBJETIVIDAD (VALOR ENTRE 0 Y 1)

PASOS EN EL ANÁLISIS DE SENTIMIENTOS

- INSTALAR LIBRERÍA TEXTBLOB: PIP INSTALL TEXTBLOB (TAMBIÉN EXISTE VADER COMO LIBRERÍA)



➤ Polaridad > 0,2 → Sentimiento positivo
➤ Polaridad [-0,2 – 0,2] → Sentimiento Neutro
➤ Polaridad < -0,2 → Sentimiento Negativo

Objetivo 0  Subjetivo 1
Subjetividad



PARTE III: RECONOCIMIENTO Y SÍNTESIS DE VOZ

DR. MARTÍN BILBAO

DIRECTOR DEL GRUPO DE INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL

DEPARTAMENTO DE INFORMÁTICA - UNPSJB

¿EN QUE CONSISTE EL RECONOCIMIENTO Y SÍNTESIS DE VOZ?

- RECONOCIMIENTO DE VOZ → TRANSCRIBIR LA VOZ O AUDIO A TEXTO



- LIBRERÍA SPEECHRECOGNITION PARA PYTHON (OTRAS.. CMU SPHINX, KALDI, WAV2LETTER++)
- SÍNTESIS DE VOZ → PROCESO INVERSO, DADO UN TEXTO EMITIR EL AUDIO CORRESPONDIENTE

PROCESO DE RECONOCIMIENTO DE VOZ

1. Crear Instancia de reconocimiento

2. Seleccionar el audio (Archivo de audio o micrófono)

3. Reconocer audio a través de API

4. Ajuste de ruido de fondo

Formatos:

- WAV
- AIFF
- AIFF-C

- ✓ Usamos la API de Google `recognize_google`
- ✓ `recognize_bing` (Microsoft)
- ✓ `recognize_google_cloud`
- ✓ `recognize_wit` (MIT)

SINTETIZADOR DE VOZ

- INSTALAR LIBRERÍA PYTTSX3

1. Crear Instancia
del sintetizador

2. Definir
parámetros
(velocidad,
frecuencia,
volumen, fuentes
etc)

3. Establecer el
idioma

4. Sintetizar texto
a voz



PARTE IV: CHATBOTS

DR. MARTÍN BILBAO

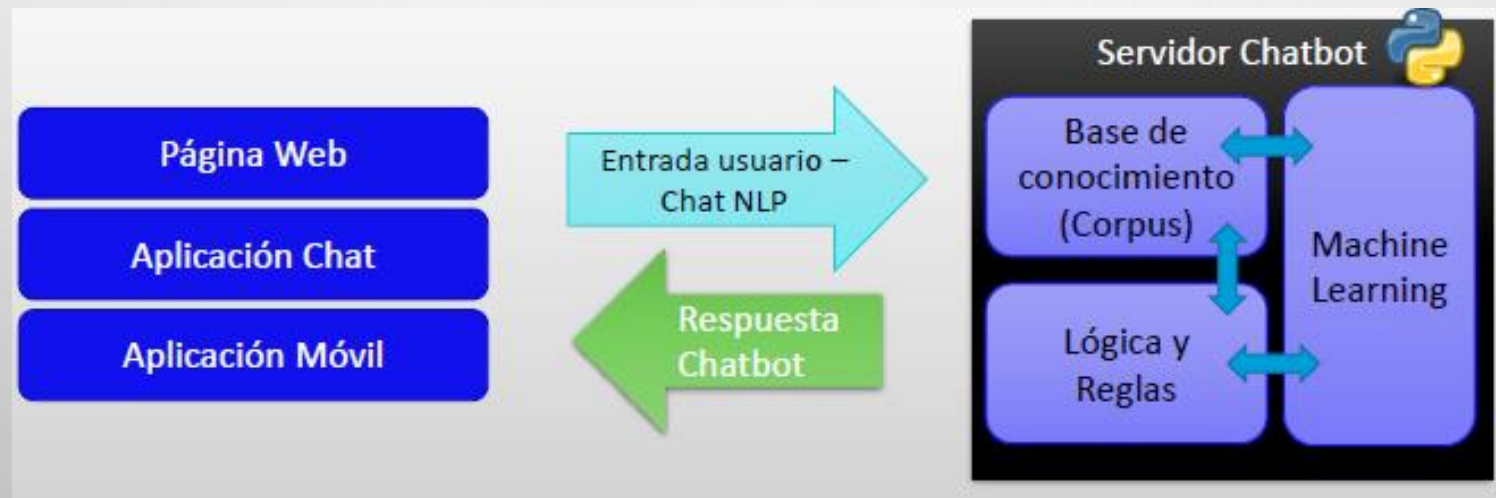
DIRECTOR DEL GRUPO DE INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL

DEPARTAMENTO DE INFORMÁTICA - UNPSJB

CHATBOTS

- ASISTENTE VIRTUAL AUTOMÁTICO QUE ASISTE EN FORMA PERSONALIZADA A CLIENTES.
- SOLUCIONAN EN TIEMPO REAL LOS PROBLEMAS MÁS COMUNES DANDO RESPUESTAS A SUS PETICIONES.
- AHORRAN TIEMPOS, REDUCEN ESFUERZOS Y COSTOS DANDO SOLUCIONES EFICIENTES.

En 2020 los chatbots manejarán el 85% de las interacciones de servicios de clientes



CREACIÓN DE UN CHATBOT

CORPUS = Base de Conocimiento (FAQs, normas, descripción de productos/servicios, soluciones técnicas,...)

1. Definición del CORPUS

2. Preprocesamiento del texto con NLTK CORPUS

2. Preprocesamiento del texto con NLTK MENSAJE USUARIO

3. Evaluar similitud Corpus - Mensaje

4. Coincidencias manuales

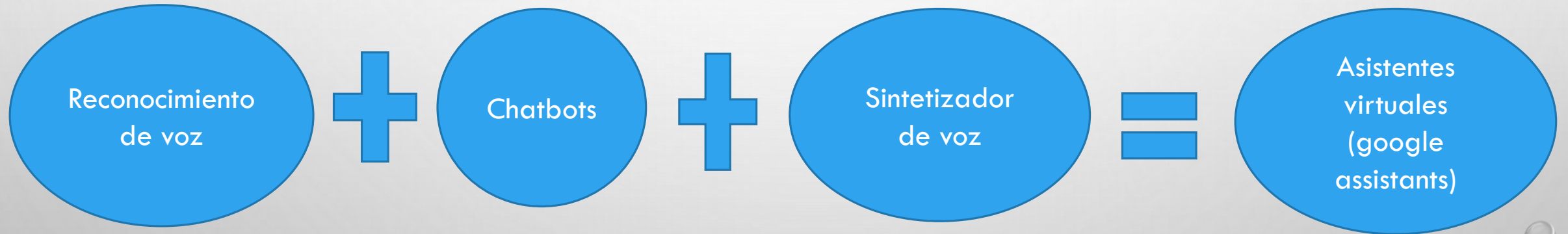
5. Generación de respuestas

Clasificación de texto

- TF-IDF
- Similitud de coseno

Saludo, agradecimiento
despedida

COMBINACIÓN DE FUNCIONES



**MUCHAS GRACIAS POR SU
ATENCIÓN**

