

AUTOR: Napoleón Alcides Pérez Arteaga

PROCESAMIENTO DE COMENTARIOS EN TIK TOK

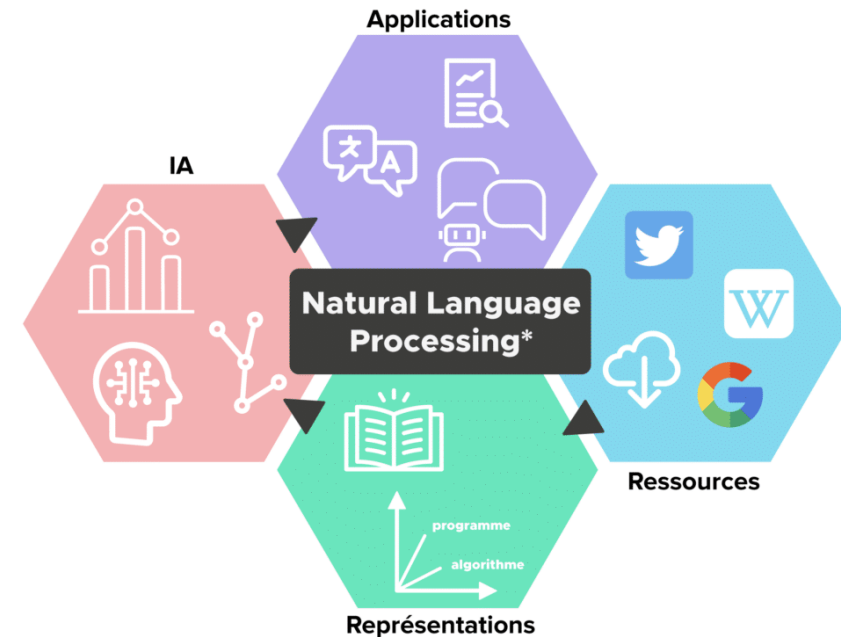
PROCESO DE OPERACIÓN DE LA SOLUCIÓN



- Identificar las palabras empleadas por cada uno de los usuarios.(Frecuencia de palabras)
- Crear un modelo de machine learning capaz de clasificar la comentarios de las publicaciones en base a su texto.
- Análisis de sentimientos, en base a librerías de publicación

COMPONENTE DE ANALÍTICA

- Clasificación automática de texto :en categorías pre-existentes ó a partir de textos completos, detectar los temas recurrentes y crear las categorías de los comentarios de usuarios de Chegg
- Modelo de inteligencia Artificial(Naive Bayes)
- El Text Mining es el proceso automático para extraer una característica específica de un texto sin una estructura clara, gracias al NPL.Al procesar esos datos de forma que las máquinas lo entiendan podemos clasificar textos y analizarlos para poder extraer el tema, sentimiento o la intención del mismo.



**Traitement Automatique du Langage Naturel*

HERRAMIENTAS DE SOFTWARE.

- NLTK: Esta es la librería con la que todos empiezan, sirve mucho para pre-procesamiento, crear los tokens, stemming, POS tagging, etc
- TextBlob: fue creada encima de NLTK y es fácil de usar. Incluye algunas funcionalidades adicionales como análisis de sentimiento y spell check.
- Gensim: contruida específicamente para modelado de temas e incluye múltiples técnicas (LDA y LSI). También calcula similitud de documentos.
- SpaCy: Puede hacer muchísimas cosas al estilo de NLTK pero es bastante más rápido.
- **Google Colab**: Desarrollo del modelo.
- Streamlit.



FASES DEL PROYECTO DE MACHINE LEARNING

- Fase 1: Entender el Problema.
- Fase 2: Definir un Criterio de Evaluación.
- Fase 3: Evaluación de la solución actual.
- Fase 4: Preparar los datos. ...
- Fase 5: Construir el modelo. ...
- Fase 6: Análisis de Errores. ...
- Fase 7: Modelo integrado en un Sistema.

INTEGRACIÓN DEL SISTEMA Y DESPLIEGUE

