



Universidad Autónoma del Estado de Hidalgo



Universidad Autónoma del Estado de Hidalgo

Instituto de Ciencias Básicas e Ingeniería

Área académica de Computación y Electrónica

Licenciatura en Ciencias Computacionales

Autómatas y Compiladores

Reporte Xpocet

Profesor:

Cornejo Veliazquez Eduardo

Desarrollado por:

Rocha Hernandez Rogelio



Título del proyecto:

“Relación de la temática y el éxito en publicaciones de Facebook mediante técnicas de Procesamiento del Lenguaje Natural”

Problema

Actualmente, muchas empresas que utilizan redes sociales, especialmente dentro del sector turístico, publican contenido digital para atraer la atención del público. Sin embargo, no existe claridad sobre qué tipos de publicaciones son más exitosas, es decir, cuáles generan más interacción como comentarios, reacciones y compartidos.

El problema principal identificado fue la falta de análisis sistemático que relacione el contenido textual (la temática) con el impacto alcanzado en Facebook.

Propuesta de Solución

El proyecto titulado “Relación de la temática y el éxito en publicaciones de Facebook mediante técnicas de Procesamiento del Lenguaje Natural” tiene como objetivo analizar publicaciones turísticas en redes sociales para identificar qué tipos de contenido generan mayor interacción, como reacciones, comentarios y compartidos.

Para lograrlo, se recopilan publicaciones de distintas páginas de turismo y se procesan mediante técnicas de *PLN*. Primero, se realiza una limpieza del texto eliminando caracteres innecesarios y normalizando palabras. Luego, se representan los textos de forma numérica utilizando métodos como TF-IDF, FastText y Sentence-BERT, que permiten entender el significado de las palabras y las relaciones entre ellas.

Posteriormente, se aplican algoritmos de reducción de dimensionalidad y clustering para agrupar las publicaciones por temática. Estos grupos se comparan con las métricas de interacción en redes sociales, con el fin de determinar qué temas y estilos de redacción generan más éxito.

El proyecto presentado en Expocet se relaciona con la materia de Autómatas y Compiladores porque emplea principios fundamentales del análisis léxico y sintáctico en el procesamiento de lenguaje natural (PLN). En esta área, los textos —como las publicaciones de Facebook— son analizados por medio de algoritmos que identifican y clasifican palabras, muy similar al funcionamiento de un compilador cuando analiza el código fuente.

Así, técnicas como la tokenización, análisis de patrones léxicos y estructuración del lenguaje se asemejan a los procesos de los autómatas finitos y los analizadores léxicos. Estas herramientas permiten reconocer secuencias válidas de palabras y comprender la estructura de los mensajes, lo cual es esencial para que el sistema agrupe y clasifique publicaciones según su contenido y éxito.

