



**Instituto de
Ciencias Básicas
e Ingeniería**

Licenciatura en Ciencias Computacionales

Autómatas y Compiladores

Profesor:

Cornejo Velazquez Eduardo

Alumno:

López López Fernando

Actividad: 3.4 Asistencia a XPOCET

Semestre: 6° **Grupo:** 3

Ciclo Escolar Julio - Diciembre 2025

1. Información del Proyecto

Título: Análisis de la Relación entre Temática y Éxito en Publicaciones de Facebook mediante Técnicas de PLN

2. Descripción del Problema y Propuesta de Solución

Problema Abordado

El proyecto busca entender la relación entre el contenido temático de las publicaciones en redes sociales (específicamente Facebook) y su éxito o engagement. Se enfrenta al desafío de analizar texto en portugués con terminología específica de gastronomía y restaurantes, donde la identificación de patrones temáticos es crucial para predecir el desempeño de las publicaciones.

Propuesta de Solución

La solución implementa un pipeline completo de análisis que incluye:

- Preparación y Enriquecimiento: Utiliza múltiples técnicas de representación textual (TF-IDF, FastText, Sentence-BERT) para capturar diferentes aspectos semánticos del contenido.
- Modelado y Clustering: Aplica 14 técnicas de reducción dimensional combinadas con 21 algoritmos de clustering, totalizando 294 experimentos para identificar patrones temáticos óptimos.
- Evaluación: Emplea métricas de validación de clusters (Silhouette, Calinski-Harabasz, Davies-Bouldin) ajustadas por cobertura de muestras para manejar ruido en los datos.

3. Aplicación de Conceptos de Autómatas y Compiladores

Unidad 1: Introducción a Compiladores

Análisis Léxico (Tokenización):

- El proyecto implementa un analizador léxico similar al de un compilador, donde las publicaciones de Facebook se descomponen en tokens individuales.
- En la tabla de "Description limpia" se observa la tokenización: "vamos tomar chopp hoje" → [vamos, tomar, chopp, hoje]
- Esto corresponde directamente a la etapa de análisis léxico de un compilador, donde el flujo de caracteres se convierte en tokens significativos.

Unidad 2: Análisis Léxico y Autómatas Finitos

Autómatas Finitos para Procesamiento de Texto:

- El proceso de identificación de términos relevantes ("caldo", "carne", "galinha") puede modelarse mediante Autómatas Finitos Deterministas (AFD).
- Los patrones de búsqueda de términos gastronómicos específicos pueden representarse como expresiones regulares procesadas por autómatas.
- La extracción de n-gramas (como "estrogroffe carne") sigue patrones que pueden ser reconocidos por autómatas finitos.

Unidad 3: Análisis Sintáctico

Gramáticas Formales para Estructura de Contenido:

- El contenido de las publicaciones podría analizarse mediante gramáticas libres de contexto para identificar estructuras recurrentes.
- Patrones como "[sustantivo] [adjetivo]" o "[ingrediente] [preparación]" siguen reglas gramaticales que pueden definirse formalmente.
- La identificación de relaciones semánticas entre términos podría beneficiarse de análisis sintáctico para entender la estructura de las frases.

4. Contribución Potencial del Curso al Proyecto

El conocimiento de Autómatas y Compiladores permitiría:

1. Optimizar el preprocesamiento mediante analizadores léxicos más eficientes
2. Validar la estructura de las publicaciones usando gramáticas formales
3. Extraer patrones complejos mediante autómatas más sofisticados
4. Organizar el vocabulario en tablas de símbolos para acceso rápido
5. Detectar errores en el formato de las publicaciones antes del análisis

5. Evidencia fotográfica

