

# Procesamiento de Reseñas

Juan Esteban Ortiz  
Jesus Israel Velez



# Introducción

- El proyecto se centra en la construcción de un programa en C para analizar reseñas de usuarios.
- Este programa mide el rendimiento de procesamiento mediante el uso de métricas de uso de CPU y memoria.
- El objetivo es evaluar la cantidad de reseñas procesadas, identificar su polaridad (positiva, negativa o neutral), y analizar la frecuencia de palabras clave asociadas



# Dataset de kaggle: Amazon Fine Food Reviews

**Este Dataset consiste en reseñas de productos alimenticios gourmet de Amazon. Las reseñas incluyen información sobre los productos y los usuarios, calificaciones y un texto plano de la reseña.**

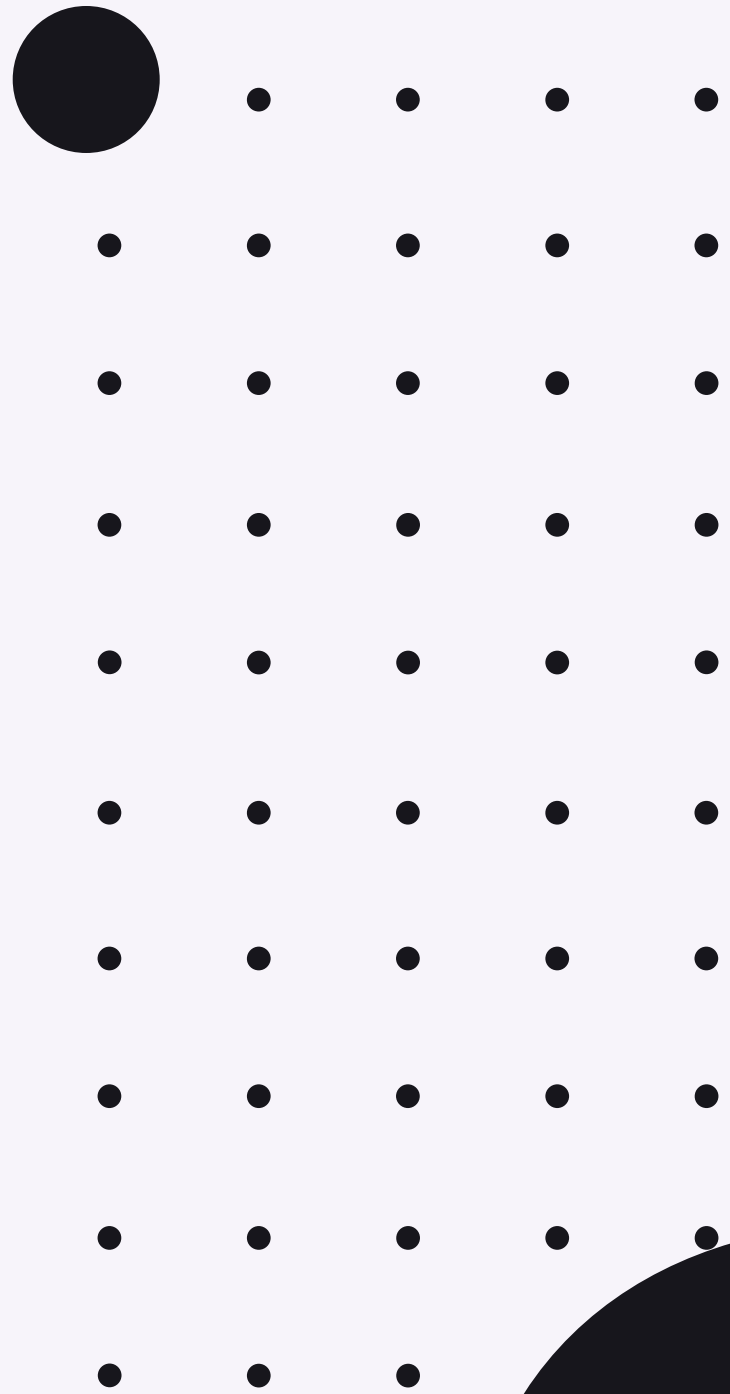
# Marco teórico

Procesamiento paralelo

Manejo de recursos compartidos

Procesamiento de texto

Medición de rendimiento



# Preprocesamiento de los datos

```
import csv

def process_csv(input_filename, output_filename):
    # Abrir el archivo CSV para leer
    with open(input_filename, 'r', encoding='utf-8') as csvfile:
        csvreader = csv.reader(csvfile)

        # Saltar la primera línea que contiene los encabezados
        next(csvreader)

        # Abrir el archivo de salida para escribir las reseñas
        with open(output_filename, 'w', encoding='utf-8') as outfile:
            count = 0 # Contador para las reseñas procesadas
            for row in csvreader:
                if count >= 10000:
                    break # Detenerse después de 1000 reseñas

                # La reseña está en la columna 'Text' (índice 9)
                review_text = row[9]

                # Escribir la reseña en el archivo de salida, seguido de una nueva línea
                outfile.write(review_text + '\n')
                count += 1

            print(f"El archivo de texto con las primeras 1000 reseñas se ha generado exitosamente: {output_filename}")

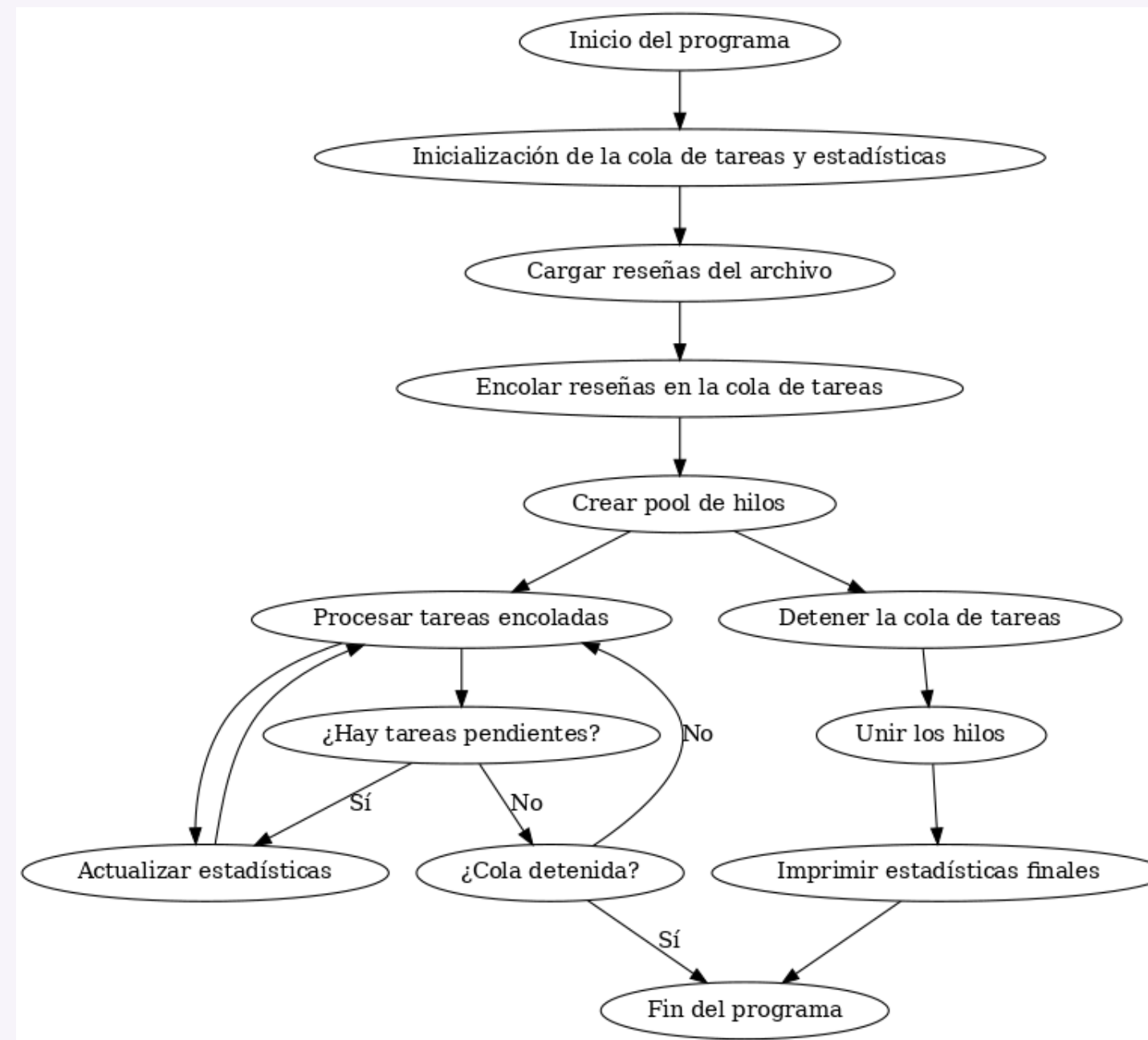
# Usar la función
input_filename = 'Reviews.csv' # Ruta del archivo CSV de entrada
output_filename = 'reviews_output.txt' # Ruta del archivo de salida

process_csv(input_filename, output_filename)
```

# Metodología

- Lectura y preprocesamiento de datos
- Procesamiento concurrente
- Sincronización
- Validación y medición de resultados

# Implementacion



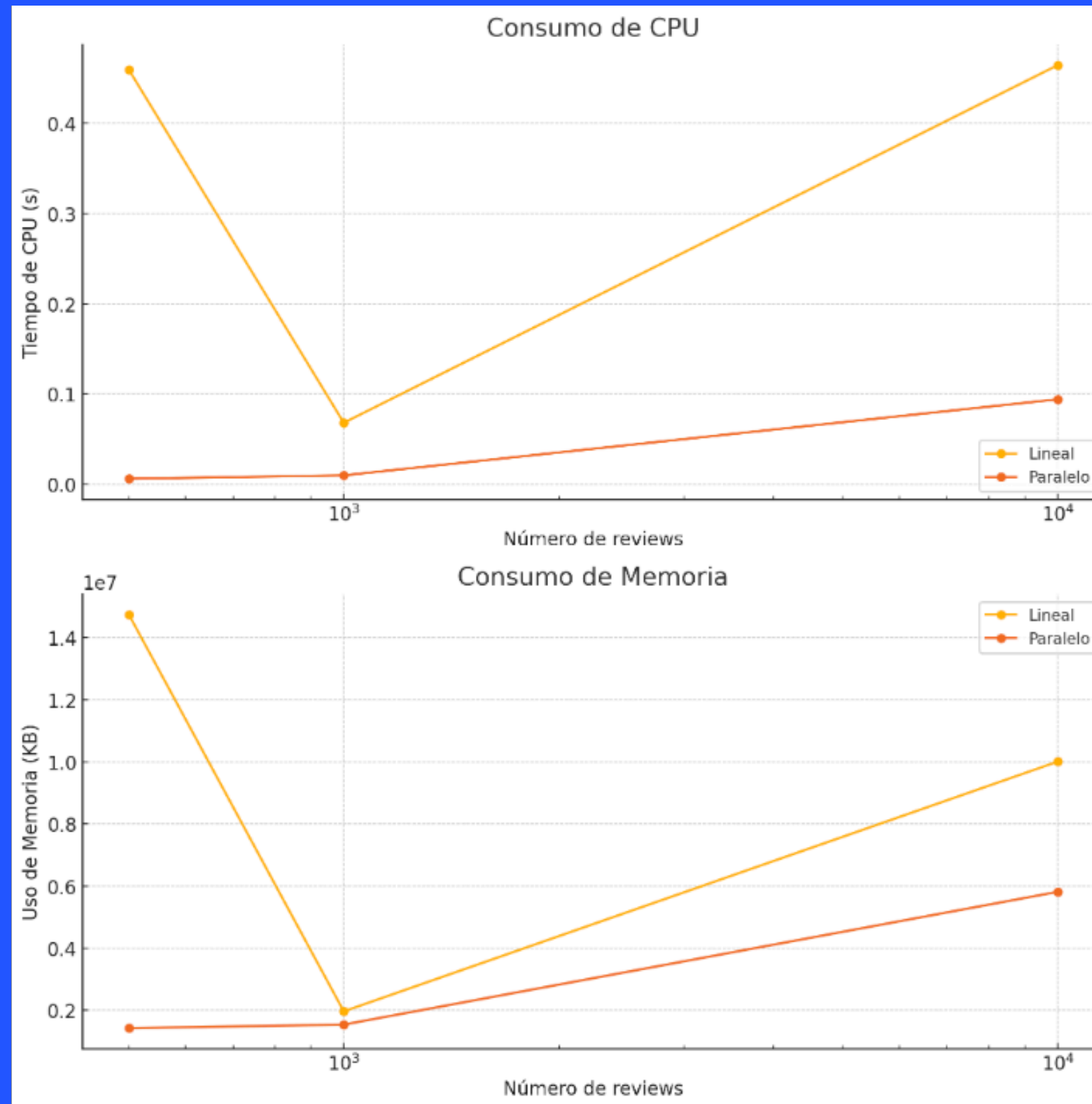


# Resultados



```
Total de reseñas: 1067  
Reseñas positivas: 615  
Reseñas negativas: 195  
Reseñas neutrales: 257  
Palabras positivas encontradas: 1753  
Palabras negativas encontradas: 805  
Tiempo de CPU usado: 0.042627 segundos  
Memoria máxima utilizada: 1507328 KB
```

# Analisis



# Conclusiones

....

- La implementación es extensible
- Las métricas de rendimiento indican que el diseño del programa es eficiente para grandes volúmenes de datos
- El uso de multithreading en el procesamiento de reseñas permite una mejora notable en el rendimiento

