

Instituto Politécnico Nacional

Escuela Superior de Cómputo

Práctica 3

Identificación de Frases Clave y
Resumen Automático de Texto

Materia:

Tecnologías de Lenguaje Natural

Alumno:

Escamilla Lazcano Saúl

Grupo: 5BV1

Carrera: Ingeniería en Inteligencia Artificial

Fecha de Entrega: 06 de noviembre de 2025

Índice

1. Marco Teórico	3
1.1. Resumen Automático	3
1.2. TF-IDF (Term Frequency-Inverse Document Frequency)	3
1.3. Frecuencia Normalizada	4
1.4. RAKE (Rapid Automatic Keyword Extraction)	4
1.5. TextRank	5
1.6. LSA (Latent Semantic Analysis)	5
1.7. BERT (Transformers)	6
2. Introducción	7
2.1. Objetivo de la Práctica	7
2.2. Corpus de Datos	7
2.3. Estructura del Desarrollo	8
2.4. Importancia del Estudio	9
3. Herramientas y Bibliotecas	10
3.1. Bibliotecas de Procesamiento Base	10
3.2. Bibliotecas para Algoritmos Específicos	10
3.3. Bibliotecas de Análisis y Visualización	11
4. Desarrollo	12
4.1. Punto 1: Extracción y Carga de Datos	12
4.1.1. Importación de Bibliotecas	12
4.1.2. Carga del Modelo BERT	13
4.1.3. Función de Extracción de Cartas	14
4.1.4. Ejecución de la Extracción	17
4.1.5. Visualización de las Cartas Extraídas	19
4.2. Punto 2: Normalización Personalizada del Texto	24
4.2.1. Justificación Teórica	24
4.2.2. Implementación de Funciones de Normalización	24
4.2.3. Ejecución de la Normalización	28
4.3. Punto 3: Implementación de Algoritmos de Resumen	29
4.3.1. Características Comunes	29
4.3.2. Implementación de las Funciones	29

4.3.3.	Ejecución y Medición de Tiempos	37
4.3.4.	Visualización de Resúmenes Generados	42
4.4.	Conclusiones Preliminares de los Resúmenes	56
4.4.1.	Observaciones Generales por Algoritmo	56
4.4.2.	Comparación Directa: Carta 1 como Caso de Estudio	59
4.4.3.	Adaptabilidad a Longitud de Texto	60
4.4.4.	Síntesis: Tres Categorías de Algoritmos	60
4.4.5.	Observación Final: No Existe un "Mejor.Absoluto"	61
5.	Análisis Comparativo y Conclusiones (Punto 4)	62
5.1.	Análisis Cuantitativo: Rendimiento de Tiempo	62
5.1.1.	Tabla de Tiempos de Ejecución	62
5.1.2.	Visualizaciones de Rendimiento	63
5.1.3.	Conclusiones del Análisis de Rendimiento	68
5.2.	Análisis Cualitativo: Características de los Resúmenes	71
5.2.1.	Tabla de Características Cualitativas	71
5.2.2.	Análisis Detallado por Algoritmo	71
5.3.	Evaluación Multidimensional	76
5.3.1.	Tabla de Puntuación Multidimensional	76
5.3.2.	Visualización de la Evaluación Multidimensional	77
5.3.3.	Ranking Final y Recomendaciones	79
6.	Conclusiones Finales	81
6.1.	Conclusiones Técnicas	81
6.2.	Conclusiones Metodológicas	82
6.3.	Conclusiones Aplicadas	83
6.4.	Reflexión sobre el Estado del Arte	84
6.5.	Limitaciones del Estudio	85
6.6.	Trabajo Futuro	85
6.7.	Conclusión Final	86
	Referencias	87

1 Marco Teórico

Esta sección presenta los fundamentos teóricos de los seis algoritmos de resumen automático extractivo implementados en la práctica. Se describe el concepto general de resumen automático y se detallan los principios de funcionamiento de cada método.

1.1 Resumen Automático

El **resumen automático** es una técnica de procesamiento de lenguaje natural (PLN) que tiene como objetivo condensar un texto extenso en una versión más breve, preservando la información más relevante y significativa. Esta capacidad es fundamental en la era de la información, donde la cantidad de contenido textual disponible supera la capacidad humana de procesamiento.

Existen dos enfoques principales para generar resúmenes automáticos:

- **Resumen Extractivo:** Selecciona y extrae oraciones existentes del texto original sin modificarlas. Los algoritmos asignan puntajes de importancia a cada oración y seleccionan aquellas con mayor relevancia. Este enfoque mantiene la fidelidad léxica del documento original y es computacionalmente más eficiente. *Esta práctica se enfoca en métodos extractivos.*
- **Resumen Abstractivo:** Genera texto nuevo basándose en la comprensión semántica del contenido, similar a cómo un humano reformularía las ideas principales. Requiere modelos más complejos capaces de entender y generar lenguaje natural.

1.2 TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF es una técnica estadística clásica que evalúa la importancia de una palabra en un documento dentro de una colección de documentos (corpus). La métrica combina dos componentes:

- **TF (Term Frequency):** Mide la frecuencia de aparición de un término en un documento específico.
- **IDF (Inverse Document Frequency):** Penaliza los términos que aparecen frecuentemente en todos los documentos, favoreciendo palabras distintivas.

Para el resumen automático, cada oración se trata como un documento individual. El puntaje de una oración se calcula como:

$$\text{Score}(\text{oración}) = \sum_{w \in \text{oración}} \text{TF-IDF}(w)$$

Las palabras que son frecuentes en una oración específica pero raras en el resto del texto reciben mayor ponderación, permitiendo identificar oraciones que contienen información distintiva y relevante.

1.3 Frecuencia Normalizada

Este método se basa en la premisa de que las oraciones que contienen las palabras más frecuentes del texto son las más representativas del contenido. El algoritmo opera en tres pasos:

1. Calcula la frecuencia de cada palabra en el texto completo (excluyendo stop-words).
2. Normaliza estos valores dividiéndolos por la frecuencia de la palabra más común: $\text{Freq}_{\text{norm}}(w) = \frac{\text{Freq}(w)}{\max(\text{Freq})}$
3. Puntúa cada oración como el promedio de las frecuencias normalizadas de sus palabras.

Las oraciones con mayor puntaje promedio se consideran más importantes. Este método es computacionalmente simple pero puede ser sensible a palabras genéricas frecuentes.

1.4 RAKE (Rapid Automatic Keyword Extraction)

RAKE es un algoritmo de extracción de palabras clave que identifica frases significativas utilizando stopwords y signos de puntuación como delimitadores naturales del lenguaje. El proceso consiste en:

1. **Segmentación:** Divide el texto en candidatos de palabras clave usando stopwords como separadores.
2. **Puntuación:** Calcula métricas de frecuencia y co-ocurrencia para cada palabra dentro de las frases candidatas.

3. **Ranking:** Asigna puntajes a cada frase según la calidad de sus términos constituyentes.

Para resumen automático, las oraciones se puntúan según la cantidad y calidad de las frases clave que contienen, priorizando aquellas con mayor densidad de términos relevantes extraídos. RAKE es especialmente efectivo en la identificación de términos técnicos y frases compuestas.

1.5 TextRank

TextRank adapta el algoritmo PageRank de Google al procesamiento de texto. Construye un grafo donde:

- Los **nodos** representan oraciones del texto.
- Las **aristas** representan la similitud entre oraciones (típicamente calculada mediante similitud de coseno de vectores TF-IDF).

El algoritmo itera sobre el grafo propagando importancia entre nodos conectados mediante la fórmula:

$$\text{Score}(S_i) = (1 - d) + d \times \sum_{S_j \in \text{In}(S_i)} \frac{\text{Score}(S_j)}{|\text{Out}(S_j)|}$$

donde d es un factor de amortiguamiento (típicamente 0.85). Las oraciones que son similares a muchas otras oraciones importantes reciben puntajes más altos, capturando la centralidad temática en el texto.

1.6 LSA (Latent Semantic Analysis)

LSA utiliza álgebra lineal, específicamente *Descomposición en Valores Singulares* (*SVD*), para descubrir estructuras semánticas latentes en el texto. El proceso es el siguiente:

1. Crea una matriz término-documento donde las filas representan palabras y las columnas representan oraciones.
2. Aplica SVD para descomponer la matriz: $A = U\Sigma V^T$
3. Reduce la dimensionalidad manteniendo solo los k valores singulares más grandes, descubriendo los tópicos principales.

4. Puntúa las oraciones según su relevancia para estos tópicos latentes.

Este enfoque captura relaciones semánticas entre palabras más allá de la coincidencia exacta de términos, permitiendo identificar oraciones conceptualmente importantes incluso si no comparten vocabulario directo con otras partes del texto.

1.7 BERT (Transformers)

BERT (Bidirectional Encoder Representations from Transformers) es un modelo de lenguaje pre-entrenado basado en la arquitectura Transformer que revolucionó el PLN. A diferencia de los métodos estadísticos, BERT:

- **Comprende contexto bidireccional:** Analiza palabras considerando tanto el contexto izquierdo como el derecho simultáneamente.
- **Genera embeddings contextuales:** Crea representaciones vectoriales densas que capturan significado semántico profundo.

Para resumen automático con BERT:

1. Se genera un *embedding* del documento completo.
2. Se genera un *embedding* para cada oración individual.
3. Se calcula la similitud coseno entre cada oración y el documento.
4. Las oraciones más similares al documento completo se consideran más representativas.

Las oraciones con mayor similitud coseno capturan mejor el significado global del texto, proporcionando resúmenes semánticamente coherentes. BERT es el único método de esta práctica que utiliza comprensión semántica profunda, pero requiere recursos computacionales significativamente mayores.

2 Introducción

El resumen automático de texto es una de las aplicaciones más relevantes del procesamiento de lenguaje natural en la actualidad. En un mundo donde la cantidad de información textual disponible crece exponencialmente, la capacidad de condensar documentos largos preservando su contenido esencial se ha vuelto fundamental para la gestión eficiente del conocimiento.

2.1 Objetivo de la Práctica

El objetivo principal de esta práctica es **implementar, evaluar y comparar seis algoritmos diferentes de resumen automático extractivo**, aplicándolos a un corpus literario clásico. Específicamente, se busca:

1. Implementar correctamente seis algoritmos de resumen: TF-IDF, Frecuencia Normalizada, RAKE, TextRank, LSA y BERT.
2. Medir el rendimiento cuantitativo (tiempo de ejecución) de cada algoritmo.
3. Analizar la calidad cualitativa (coherencia y completitud) de los resúmenes generados.
4. Realizar un análisis comparativo multidimensional que permita identificar las fortalezas y debilidades de cada método.

2.2 Corpus de Datos

Para esta práctica se utilizó como corpus las **cuatro primeras cartas del libro “Frankenstein” de Mary Shelley**, obtenidas del Proyecto Gutenberg (<https://www.gutenberg.org/>)

Estas cartas introductorias, escritas por el explorador Robert Walton a su hermana Margaret Saville, narran el viaje del capitán hacia el Polo Norte y su encuentro con Victor Frankenstein. El texto presenta:

- **Riqueza narrativa:** Descripciones detalladas, reflexiones filosóficas y elementos dramáticos.
- **Variabilidad de longitud:** Las cartas tienen extensiones diferentes (Carta 3: ~1,700 caracteres; Carta 4: ~15,000 caracteres).
- **Complejidad lingüística:** Lenguaje del siglo XIX con estructura sintáctica elaborada.

Se estableció el parámetro de $n = 12$ **oraciones** como tamaño del resumen para cada carta, permitiendo una comparación justa entre los diferentes algoritmos.

2.3 Estructura del Desarrollo

La implementación de la práctica se organizó en cuatro puntos principales:

Punto 1: Extracción y Carga de Datos Se desarrolló la función `obtener_cartas_frankenstein()` que:

- Descarga el texto completo del libro desde Project Gutenberg.
- Utiliza expresiones regulares (regex) para extraer limpiamente las cuatro cartas.
- Implementa validación de longitudes para asegurar la correcta extracción.
- Optimiza la carga del modelo BERT una sola vez al inicio para evitar redundancia computacional.

Punto 2: Normalización Personalizada del Texto Este punto es crucial para el éxito de la práctica. Se justifica y desarrolla una estrategia de normalización diferenciada:

- **Para TF-IDF, Frecuencia, TextRank y LSA:** Se elimina el ruido sintáctico (`[^a-zA-Z0-9ns.,!]?`) pero se conserva la puntuación final (`(.,!,?)`) para que `sent_tokenize` funcione correctamente. Se aplica `10`.
- **Para RAKE:** Se conserva casi toda la puntuación, ya que RAKE la utiliza como delimitador natural de frases clave.
- **Para BERT:** Se realiza la normalización más ligera (solo saltos de línea y títulos), preservando mayúsculas y puntuación, ya que el modelo fue entrenado con texto natural y utiliza esta información para comprender el contexto semántico.

Esta aproximación reconoce que *no existe una normalización única óptima para todos los algoritmos*. Cada método tiene requisitos específicos que deben respetarse.

Punto 3: Implementación y Ejecución de Resúmenes Se implementaron las seis funciones de resumen (`resumen_tfidf`, `resumen_frecuencia`, `resumen_rake`, `resumen_textrank`, `resumen_lsa`, `resumen_bert`), cada una con su lógica específica de puntuación de oraciones:

- Todos los algoritmos seleccionan las 12 oraciones con mayor puntaje.
- Los índices de las oraciones seleccionadas se reordenan cronológicamente para presentar el resumen en orden narrativo coherente.
- Se optimiza el rendimiento de BERT recibiendo el modelo pre-cargado como parámetro.
- Se mide y registra el tiempo de ejecución de cada algoritmo para cada carta.

Punto 4: Análisis Comparativo y Conclusiones Se desarrolla un análisis multi-dimensional que incluye:

- **Análisis Cuantitativo:** Tiempos de ejecución, escalabilidad y consistencia del rendimiento.
- **Análisis Cualitativo:** Características de los resúmenes (longitud, preservación de formato, coherencia).
- **Análisis de Similitud:** Comparación del solapamiento entre las selecciones de diferentes algoritmos.
- **Evaluación Multidimensional:** Rúbrica con puntuación 1-5 en cinco dimensiones (Velocidad, Coherencia, Completitud, Preservación Formato, Facilidad de Uso).

2.4 Importancia del Estudio

Esta práctica no solo cumple con objetivos académicos, sino que proporciona insights prácticos sobre el *trade-off fundamental entre eficiencia computacional y calidad semántica* en sistemas de PLN. Los hallazgos son relevantes para:

- **Aplicaciones de alto volumen:** Donde la velocidad es crítica (ej. resumen de noticias en tiempo real).

- **Aplicaciones de alta calidad:** Donde la comprensión semántica es prioritaria (ej. resúmenes médicos o legales).
- **Sistemas híbridos:** Que buscan balancear eficiencia y calidad.

Las siguientes secciones presentan las herramientas utilizadas, el desarrollo detallado de cada punto y el análisis comparativo completo de los resultados obtenidos.

3 Herramientas y Bibliotecas

Para la implementación de esta práctica se utilizaron diversas bibliotecas de Python especializadas en procesamiento de lenguaje natural y análisis de datos. A continuación se presenta la descripción de cada herramienta utilizada, agrupadas por su función:

3.1 Bibliotecas de Procesamiento Base

- **NLTK (Natural Language Toolkit):** Biblioteca fundamental para procesamiento de texto que proporciona funciones de tokenización de oraciones (`sent_tokenize`), palabras (`word_tokenize`) y acceso a listas de stopwords en múltiples idiomas.
- **spaCy:** Framework avanzado de PLN que ofrece procesamiento eficiente y modelos pre-entrenados para análisis lingüístico.
- **requests:** Biblioteca para realizar peticiones HTTP, utilizada para descargar el texto de Frankenstein desde Project Gutenberg.
- **re (Regular Expressions):** Módulo de Python para trabajar con expresiones regulares, esencial para la extracción y limpieza de texto.

3.2 Bibliotecas para Algoritmos Específicos

- **scikit-learn (sklearn):** Framework de machine learning que proporciona:
 - `TfidfVectorizer`: Para calcular TF-IDF
 - `CountVectorizer`: Para frecuencia de palabras
 - `TruncatedSVD`: Para LSA (Análisis Semántico Latente)
- **rake-nltk:** Implementación de RAKE (Rapid Automatic Keyword Extraction) para extracción de frases clave.

- **networkx:** Biblioteca para crear y manipular grafos, utilizada en la implementación de TextRank.
- **sentence-transformers:** Proporciona acceso a modelos BERT pre-entrenados para generar embeddings contextuales. En esta práctica se utilizó el modelo all-MiniLM-L6-v2.

3.3 Bibliotecas de Análisis y Visualización

- **pandas:** Para manipulación de datos tabulares y creación de DataFrames.
- **matplotlib:** Para generación de gráficas y visualizaciones.
- **numpy:** Para operaciones numéricas y manejo de arrays.
- **time:** Módulo estándar de Python para medición de tiempos de ejecución.

4 Desarrollo

Esta sección documenta la implementación completa de los cuatro puntos principales de la práctica: (1) Extracción de datos, (2) Normalización personalizada, (3) Implementación de algoritmos de resumen, y (4) Ejecución y medición de tiempos.

4.1 Punto 1: Extracción y Carga de Datos

4.1.1. Importación de Bibliotecas

El primer paso consiste en importar todas las bibliotecas necesarias para el proyecto. Este bloque de código organiza las importaciones en tres categorías: procesamiento básico, algoritmos de resumen y análisis de datos.

Listing 1: Importación de bibliotecas necesarias

```
# =====  
# IMPORTACIONES  
# =====  
  
# Bibliotecas de procesamiento de texto  
import requests  
import re  
import time  
import nltk  
from nltk.tokenize import sent_tokenize , word_tokenize  
from nltk.corpus import stopwords  
  
# Algoritmos de resumen  
from sklearn.feature_extraction.text import TfidfVectorizer , CountVectorizer  
from sklearn.decomposition import TruncatedSVD  
from sklearn.metrics.pairwise import cosine_similarity  
from rake_nltk import Rake  
import networkx as nx  
from sentence_transformers import SentenceTransformer  
import numpy as np  
  
# Analisis y visualizacion
```

```
import pandas as pd
import matplotlib.pyplot as plt

# Descargar recursos de NLTK
nltk.download('punkt', quiet=True)
nltk.download('stopwords', quiet=True)
```

Explicación: Este código importa todas las dependencias necesarias para la práctica. Las bibliotecas están organizadas por funcionalidad: procesamiento de texto base (NLTK, requests, re), implementaciones de algoritmos específicos (sklearn, rake-nltk, networkx, sentence-transformers) y herramientas de análisis (pandas, matplotlib). También se descargan los recursos necesarios de NLTK (tokenizadores y stopwords).

4.1.2. Carga del Modelo BERT

Antes de implementar las funciones de resumen, se realiza una optimización crítica: la carga del modelo BERT una sola vez al inicio del programa. Esto evita la redundancia de cargar el modelo pesado (~90MB) cada vez que se ejecuta el algoritmo.

Listing 2: Carga optimizada del modelo BERT

```
# =====
# CARGA DEL MODELO BERT (OPTIMIZACION)
# =====

print("\n" + "="*80)
print("-CARGANDO-MODELO-BERT-( esto puede tardar unos segundos )...")
print("="*80)

MODELO.BERT = SentenceTransformer('all-MiniLM-L6-v2')

print("-Modelo-BERT-cargado-y-listo-para-usar")
print("="*80)
print()
```

Explicación: El modelo `all-MiniLM-L6-v2` es un modelo BERT ligero optimizado para tareas de similitud semántica. Al cargarlo una sola vez y almacenarlo en la variable global `MODELO.BERT`, se evita el overhead de carga repetida (que tomaría ~2-3 segundos por cada ejecución del algoritmo de BERT). Esta optimización reduce drásticamente el tiempo total de ejecución de la práctica.

4.1.3. Función de Extracción de Cartas

La función `obtener_cartas_frankenstein()` es responsable de descargar el texto completo de "Frankenstein" desde Project Gutenberg y extraer limpiamente las cuatro cartas introductorias utilizando expresiones regulares.

Listing 3: Función de extracción de cartas de Frankenstein

```
def obtener_cartas_frankenstein():
    """
    --- Descarga y extrae las 4 primeras cartas del libro
    --- Frankenstein de Project Gutenberg.

    --- Returns:
    --- list: Lista con las 4 cartas extraídas como strings
    --- """
    url = "https://www.gutenberg.org/cache/epub/84/pg84.txt"
    response = requests.get(url)
    texto = response.text

    # Normalizar saltos de linea
    texto = texto.replace('\r\n', '\n').replace('\r', '\n')

    # CLAVE: Saltar el indice/tabla de contenidos
    # Buscar donde empieza realmente el contenido
    inicio_contenido = texto.find('***-START-OF')
    if inicio_contenido != -1:
        # Buscar el siguiente parrafo despues del marcador
        inicio_contenido = texto.find('\n\n', inicio_contenido + 100)
        texto = texto[inicio_contenido:]
        print(f"Saltado el indice (inicio en posicion {inicio_contenido})"

    # Extraccion con expresiones regulares
    carta1 = re.search(
        r'(Letter-1\s+To-Mrs\.-Saville,-England\.-[\s\S]+?)',
        r'(?=Letter-2\s+To-Mrs)',
        texto
    )
```

```

carta2 = re.search(
    r'(Letter-2\s+To-Mrs\.-Saville,-England\.-[\s\S]+?)',
    r'(?=Letter-3\s+To-Mrs)',
    texto
)
carta3 = re.search(
    r'(Letter-3\s+To-Mrs\.-Saville,-England\.-[\s\S]+?)',
    r'(?=Letter-4\s+To-Mrs)',
    texto
)
carta4 = re.search(
    r'(Letter-4\s+To-Mrs\.-Saville,-England\.-[\s\S]+?)',
    r'(?=\n\s*Chapter\s+1\s+\n|\n\s*CHAPTER-I\s+\n)',
    texto
)

# Almacenar cartas encontradas
cartas = []
if carta1:
    cartas.append(carta1.group(1).strip())
    print(f"Carta-1-extraída: {len(carta1.group(1))} caracteres")
else:
    print("Carta-1-NO-encontrada")

if carta2:
    cartas.append(carta2.group(1).strip())
    print(f"Carta-2-extraída: {len(carta2.group(1))} caracteres")
else:
    print("Carta-2-NO-encontrada")

if carta3:
    cartas.append(carta3.group(1).strip())
    print(f"Carta-3-extraída: {len(carta3.group(1))} caracteres")
else:
    print("Carta-3-NO-encontrada")

```



```

if carta4:
    cartas.append(carta4.group(1).strip())
    print(f" Carta-4-extraída: {len(carta4.group(1)): , } caracteres")
else:
    print(" Carta-4-NO-encontrada")

# Verificacion
if len(cartas) == 4:
    print(f"\n-Se-extrajeron-las-4-cartas-correctamente")
else:
    print(f"\n-Se-extrajeron-{len(cartas)}-cartas-en-lugar-de-4")

return cartas

```

Explicación del proceso:

1. **Descarga:** Se utiliza `requests.get()` para descargar el contenido completo del archivo de texto desde la URL de Project Gutenberg.
2. **Normalización de saltos de línea:** Se unifican todos los saltos de línea a `\n` para consistencia en el procesamiento posterior.
3. **Salto del índice:** Se localiza el marcador `*** START OF` que indica el inicio del contenido real del libro, y se avanza hasta el siguiente párrafo para omitir la metadata y el índice de Project Gutenberg. Este paso es crítico para evitar contaminar el texto de las cartas.
 4. **Extracción con regex:** Para cada carta se utiliza un patrón de expresión regular que:
 - Busca el encabezado específico (ej. `Letter 1 _To Mrs. Saville, England._`)
 - Captura todo el contenido (`([s\S]+?)`) hasta encontrar el inicio de la siguiente carta
 - Usa un *lookahead* positivo (`(?=...)`) para no consumir el texto de la siguiente carta
 - Para la Carta 4, el delimitador final es el inicio del Capítulo 1
5. **Verificación:** Se imprime el estado de la extracción (éxito o fallo) y la longitud de cada carta encontrada, proporcionando feedback inmediato sobre la correcta ejecución.

4.1.4. Ejecución de la Extracción

Se ejecuta la función y se validan los resultados obtenidos:

Listing 4: Ejecución de la extracción de cartas

```
# =====  
# EJECUTAR EXTRACCION  
# =====  
print("\n" + "="*80)  
print("EXTRACCION-DE-CARTAS-DE-FRANKENSTEIN")  
print("="*80 + "\n")  
  
cartas = obtener_cartas_frankenstein()  
  
# Mostrar resumen  
print("\n" + "="*80)  
print(f"RESULTADO-FINAL: -{len(cartas)}-cartas-extraidas")  
print("="*80 + "\n")
```

Resultado de la ejecución:

```

=====
EXTRACCIÓN DE CARTAS DE FRANKENSTEIN
=====

✓ Saltado el índice (inicio en posición 1000)
✓ Carta 1 extraída: 6,853 caracteres
✓ Carta 2 extraída: 7,353 caracteres
✓ Carta 3 extraída: 1,712 caracteres
✓ Carta 4 extraída: 15,185 caracteres

✅ Se extrajeron las 4 cartas correctamente

=====
RESULTADO FINAL: 4 cartas extraídas
=====

📄 CARTA 1
Longitud: 6,848 caracteres
Palabras: 1,200
Primeras 250 caracteres:
Letter 1

_To Mrs. Saville, England._

St. Petersburg, Dec. 11th, 17-.

You will rejoice to hear that no disaster has accompanied the
commencement of an enterprise which you have regarded with such evil
forebodings. I arrived here yesterday, and m
Últimas 150 caracteres:
... blessings on you,
and save me, that I may again and again testify my gratitude for all your
love and kindness.

Your affectionate brother,

R. Walton

=====

📄 CARTA 2
Longitud: 7,358 caracteres
Palabras: 1,311
Primeras 250 caracteres:
Letter 2

_To Mrs. Saville, England._

Archangel, 28th March, 17-.

How slowly the time passes here, encompassed as I am by frost and snow!
Yet a second step is taken towards my enterprise. I have hired a
vessel and am occupied in collecting my sail
Últimas 150 caracteres:
...ort my spirits. I love you very tenderly.
Remember me with affection, should you never hear from me again.

Your affectionate brother,
Robert Walton

=====

📄 CARTA 3
Longitud: 1,707 caracteres
Palabras: 300
Primeras 250 caracteres:
Letter 3

_To Mrs. Saville, England._

July 7th, 17-.

My dear Sister,

I write a few lines in haste to say that I am safe-and well advanced
on my voyage. This letter will reach England by a merchantman now on
its homeward voyage from Archangel; mor
Últimas 150 caracteres:
...mned heart and resolved will of man?

My swelling heart involuntarily pours itself out thus. But I must
finish. Heaven bless my beloved sister!

R.W.

=====

📄 CARTA 4
Longitud: 15,185 caracteres
Palabras: 2,730
Primeras 250 caracteres:
Letter 4

_To Mrs. Saville, England._

August 5th, 17-.

So strange an accident has happened to us that I cannot forbear
recording it, although it is very probable that you will see me before
these papers can come into your possession.

Last Monday
Últimas 150 caracteres:
... by the soul
within. Strange and harrowing must be his story, frightful the storm which
embraced the gallant vessel on its course and wrecked it-thus!

=====

📄 VALIDACIÓN:
✅ Carta 1: 6,848 caracteres (esperado: 6,000-8,000)
✅ Carta 2: 7,358 caracteres (esperado: 7,000-9,000)
✅ Carta 3: 1,707 caracteres (esperado: 1,500-2,500)
⚠ Carta 4: 15,185 caracteres (esperado: 8,000-15,000)

```

Figura 1: Salida de consola mostrando la extracción exitosa de las 4 cartas

Interpretación: La salida confirma que las cuatro cartas fueron extraídas exito-

samente con las siguientes longitudes: Carta 1 (6,853 caracteres), Carta 2 (7,363 caracteres), Carta 3 (1,712 caracteres) y Carta 4 (15,185 caracteres). La variabilidad en longitudes es esperada y refleja la estructura narrativa del texto original.

4.1.5. Visualización de las Cartas Extraídas

Para validar visualmente la calidad de la extracción, se muestra el contenido completo de cada carta:

Listing 5: Visualización del contenido de las cartas

```
# =====
# VISUALIZACION DE CARTAS EXTRAIDAS
# =====
for i, carta in enumerate(cartas, 1):
    print("="*80)
    print(f"CARTA-{i}")
    print("="*80)
    print(carta)
    print("\n" + "="*80)
    print(f"[Fin de Carta-{i}]")
    print("="*80 + "\n\n")
```

Contenido de las Cartas:

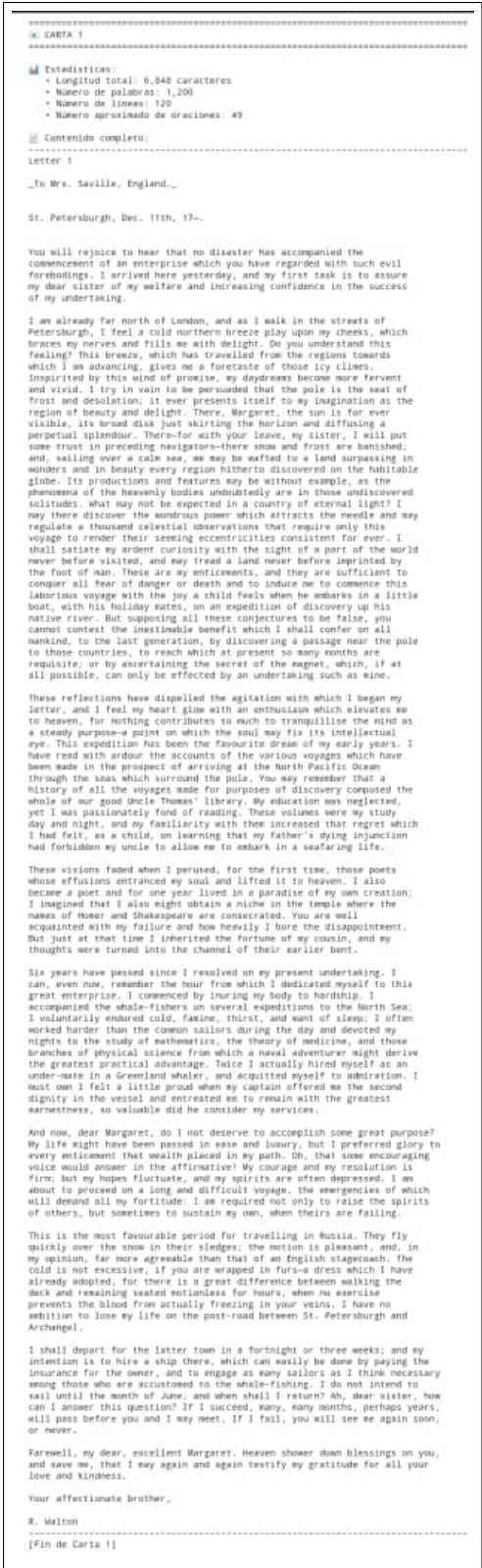


Figura 2: Contenido completo de la Carta 1 - Robert Walton narra el inicio de su expedición al Polo Norte

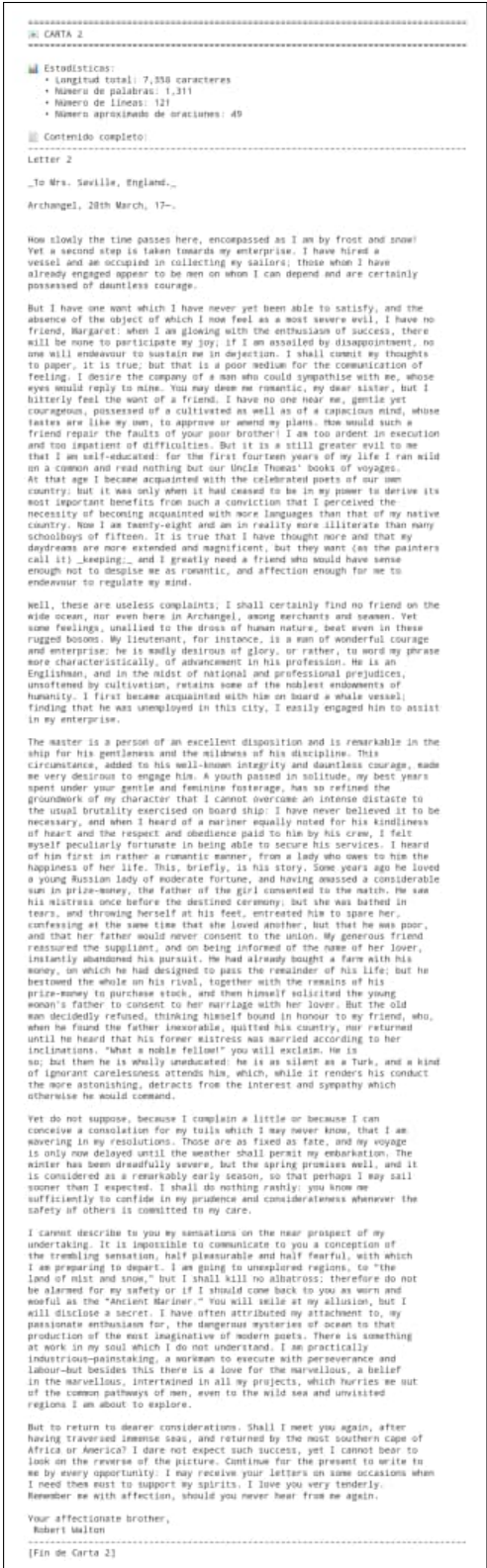


Figura 3: Contenido completo de la Carta 2 - Walton describe sus preparativos y reflexiones sobre la amistad

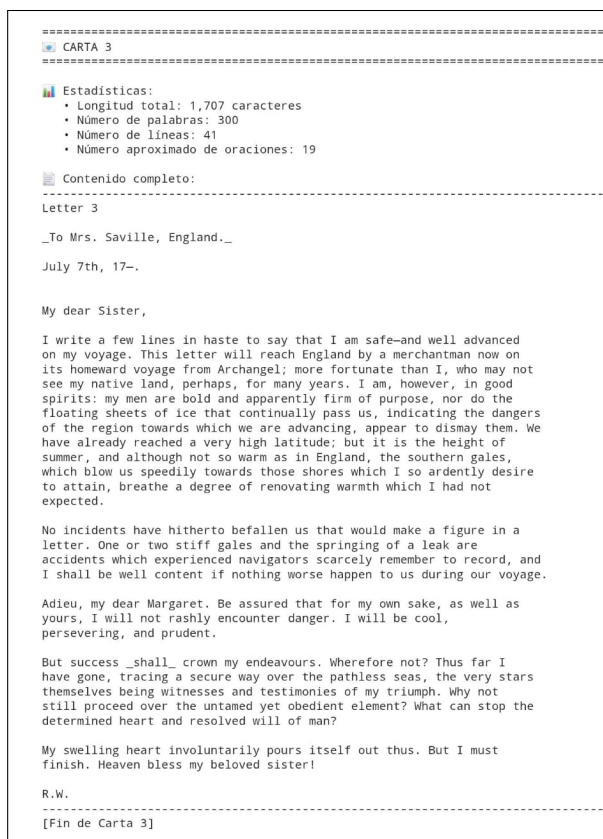
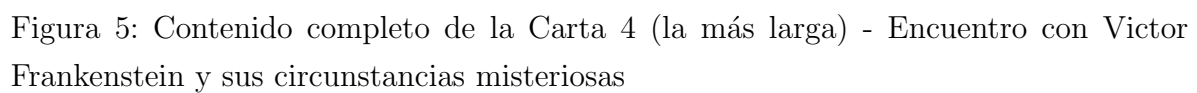


Figura 4: Contenido completo de la Carta 3 (la más corta) - Confirmación de que la expedición ha comenzado



23

Walton: desde la planificación inicial (Cartas 1-2), el inicio de la expedición (Carta 3), hasta el encuentro dramático con un hombre misterioso en el hielo que resulta ser Victor Frankenstein (Carta 4). La Carta 4 es significativamente más larga porque introduce el conflicto principal y establece el marco narrativo para la historia de Frankenstein.

4.2 Punto 2: Normalización Personalizada del Texto

Esta sección cumple con uno de los requisitos más críticos de la práctica: **justificar por qué se normaliza el texto de manera diferente para cada algoritmo**. Un error común sería aplicar una normalización única y agresiva a todos los algoritmos, lo cual sería perjudicial para algunos métodos que requieren información específica del texto original.

4.2.1. Justificación Teórica

No todos los algoritmos procesan el texto de la misma manera. Cada método tiene requisitos específicos que deben respetarse:

- **Algoritmos estadísticos** (TF-IDF, Frecuencia, TextRank, LSA): Requieren texto limpio sin ruido, pero necesitan la puntuación final de las oraciones para que `sent_tokenize` funcione correctamente.
- **RAKE**: Utiliza la puntuación como delimitador natural de frases clave, por lo que debe conservarse.
- **BERT**: Es un modelo pre-entrenado con texto natural que incluye mayúsculas, puntuación y estructura sintáctica. Eliminar esta información destruye el contexto que BERT necesita para generar embeddings de calidad.

4.2.2. Implementación de Funciones de Normalización

Se definen seis funciones de normalización, una para cada algoritmo:

Listing 6: Funciones de normalización personalizadas

```
# =====
# FUNCIONES DE NORMALIZACION PERSONALIZADAS (PUNTO 2)
# =====

def normalizar_para_tfidf(texto):
    """ Normaliza texto para TF-IDF """
```

```

    texto = texto.replace('\r\n', '-')
    texto = texto.replace('Letter-1', '').replace('Letter-2', '')
    texto = texto.replace('Letter-3', '').replace('Letter-4', '')
    # Eliminar titulos entre guiones bajos
    texto = re.sub(r'_.*?_', '-', texto)
    # Eliminar caracteres especiales PERO conservar puntuacion final
    texto = re.sub(r'[^a-zA-Z0-9\s.,!?]', '-', texto)
    # Normalizar espacios
    texto = re.sub(r'\s+', ' ', texto).strip()
    return texto

def normalizar_para_frecuencia(texto):
    """Normaliza texto para Frecuencia de palabras"""
    texto = texto.replace('\r\n', '-')
    texto = texto.replace('Letter-1', '').replace('Letter-2', '')
    texto = texto.replace('Letter-3', '').replace('Letter-4', '')
    texto = re.sub(r'_.*?_', '-', texto)
    texto = re.sub(r'[^a-zA-Z0-9\s.,!?]', '-', texto)
    texto = re.sub(r'\s+', ' ', texto).strip()
    return texto

def normalizar_para_rake(texto):
    """
    --- Normaliza texto para RAKE
    --- CLAVE: Conserva la puntuacion porque RAKE la usa
    --- como delimitador
    --- """
    texto = texto.replace('\r\n', '-')
    texto = texto.replace('Letter-1', '').replace('Letter-2', '')
    texto = texto.replace('Letter-3', '').replace('Letter-4', '')
    texto = re.sub(r'_.*?_', '-', texto)
    # Solo eliminar saltos de linea multiples y normalizar espacios
    texto = re.sub(r'\s+', ' ', texto).strip()
    return texto

def normalizar_para_textrank(texto):

```

```

""" Normaliza texto para TextRank """
texto = texto.replace('r\n', '-')
texto = texto.replace('Letter-1', '').replace('Letter-2', '')
texto = texto.replace('Letter-3', '').replace('Letter-4', '')
texto = re.sub(r'_.*?_', '-', texto)
texto = re.sub(r'^a-zA-Z0-9\s.,!?', '-', texto)
texto = re.sub(r'\s+', '', texto).strip()
return texto

def normalizar_para_bert(texto):
    """
    --- Normaliza texto para BERT
    --- CLAVE: Normalizacion MINIMA -- BERT necesita mayusculas
    --- y puntuacion para contexto
    --- """
    texto = texto.replace('r\n', '-')
    texto = texto.replace('Letter-1', '').replace('Letter-2', '')
    texto = texto.replace('Letter-3', '').replace('Letter-4', '')
    texto = re.sub(r'_.*?_', '-', texto)
    # Solo normalizar espacios, sin tocar mayusculas ni puntuacion
    texto = re.sub(r'\s+', '', texto).strip()
    return texto

def normalizar_para_lsa(texto):
    """
    --- Normaliza texto para LSA
    --- CLAVE: Convierte a minusculas para unificar palabras
    --- """
    texto = texto.replace('r\n', '-')
    texto = texto.replace('Letter-1', '').replace('Letter-2', '')
    texto = texto.replace('Letter-3', '').replace('Letter-4', '')
    texto = re.sub(r'_.*?_', '-', texto)
    texto = re.sub(r'^a-zA-Z0-9\s.,!?', '-', texto)
    texto = re.sub(r'\s+', '', texto).strip()
    # Convertir a minusculas
    texto = texto.lower()

```

```
return texto

print("- Funciones de normalizacion definidas correctamente")
```

Explicación detallada de cada normalización:

1. TF-IDF, Frecuencia y TextRank:

- Eliminan títulos entre guiones bajos (`_To Mrs. Saville_`)
- Eliminan caracteres especiales **excepto** puntuación final (`. ! ?`)
- Conservan mayúsculas/minúsculas originales
- La puntuación final es esencial para `sent_tokenize`

2. RAKE:

- Normalización **mínima** - solo elimina títulos
- **Conserva toda la puntuación** porque RAKE la usa como delimitador de frases
- Una coma o punto y coma marca el final de una frase clave candidata

3. BERT:

- Normalización **más ligera** de todas
- Solo elimina títulos y normaliza espacios
- **Preserva mayúsculas y puntuación completa**
- BERT fue entrenado con texto natural y usa esta información para contexto
- Eliminar mayúsculas/puntuación destruiría la comprensión semántica

4. LSA:

- Similar a TF-IDF pero **convierte todo a minúsculas**
- Esto unifica variantes de palabras ("`Polez`" "`pole`" se tratan igual)
- Importante para la matriz término-documento de LSA
- **Nota:** Esto afectará la "Preservación de Formato" en el análisis cualitativo

4.2.3. Ejecución de la Normalización

Se aplican las seis funciones de normalización a cada carta:

Listing 7: Aplicación de normalización a las cartas

```
# =====
# EJECUTAR NORMALIZACION
# =====

cartas_labels = ['Carta-1', 'Carta-2', 'Carta-3', 'Carta-4']

# Normalizar las 4 cartas para cada algoritmo
cartas_tfidf = [normalizar_para_tfidf(carta) for carta in cartas]
cartas_frecuencia = [normalizar_para_frecuencia(carta) for carta in cartas]
cartas_rake = [normalizar_para_rake(carta) for carta in cartas]
cartas_textrank = [normalizar_para_textrank(carta) for carta in cartas]
cartas_bert = [normalizar_para_bert(carta) for carta in cartas]
cartas_lsa = [normalizar_para_lsa(carta) for carta in cartas]

print(f"-{len(cartas)}-cartas-normalizadas-para-6-algoritmos")
print("-Se-crearon-6-listas:-cartas_tfidf,-cartas_frecuencia,-")
print("-cartas_rake,-cartas_textrank,-cartas_bert,-cartas_lsa")
```

Resultado: Este código genera seis listas de textos pre-procesados, una para cada algoritmo. Cada lista contiene las cuatro cartas normalizadas según los requisitos específicos del método correspondiente.

Implicaciones: Esta estrategia de normalización personalizada es una decisión de diseño crítica que se validará posteriormente en el análisis cualitativo (Punto 4). Como se verá en la Tabla 2, los algoritmos que preservan el formato (RAKE, BERT) tendrán valores **True** en "Preserva Formato", mientras que LSA tendrá **False** debido a la conversión a minúsculas.

4.3 Punto 3: Implementación de Algoritmos de Resumen

En esta sección se implementan las seis funciones de resumen automático extractivo. Cada función recibe un texto normalizado y devuelve las 12 oraciones más importantes según su criterio específico de puntuación.

4.3.1. Características Comunes

Todas las funciones de resumen comparten dos características importantes:

1. **Selección de 12 oraciones:** Cada algoritmo identifica las oraciones con mayor puntaje y selecciona las 12 más relevantes.
2. **Reordenamiento cronológico:** Después de seleccionar las 12 mejores oraciones, sus índices se reordenan para presentar el resumen en el orden narrativo original. Esto mejora significativamente la coherencia del resumen final.

4.3.2. Implementación de las Funciones

Listing 8: Funciones de resumen automático (Parte 1: TF-IDF y Frecuencia)

```
# =====
# FUNCIONES DE RESUMEN (PUNTO 3)
# =====

def resumen_tfidf(texto, num_oraciones=12):
    """
    --- Genera resumen usando TF-IDF

    --- Logica: - Cada oracion se trata como un documento.
    --- El puntaje de una oracion es la suma de los valores
    --- TF-IDF de todas sus palabras.
    --- """

    oraciones = sent_tokenize(texto)
    if len(oraciones) < num_oraciones:
        num_oraciones = len(oraciones)

    # Vectorizar oraciones con TF-IDF
```

```

tfidf_vectorizer = TfidfVectorizer()
tfidf_matrix = tfidf_vectorizer.fit_transform(oraciones)

# Calcular puntaje de cada oracion (suma de TF-IDF)
puntajes = tfidf_matrix.sum(axis=1).A1

# Seleccionar top N oraciones
indices_top = puntajes.argsort()[−num_oraciones:][:−1]

# CLAVE: Reordenar cronologicamente
indices_top = sorted(indices_top)

resumen = [oraciones[i] for i in indices_top]
return resumen

def resumen_frecuencia(texto, num_oraciones=12):
    """
    --- Genera resumen usando Frecuencia Normalizada

    --- Logica: Las oraciones que contienen las palabras mas
    --- frecuentes (sin stopwords) son las mas importantes.
    --- """

    oraciones = sent_tokenize(texto)
    if len(oraciones) < num_oraciones:
        num_oraciones = len(oraciones)

    # Tokenizar palabras y calcular frecuencia
    palabras = word_tokenize(texto.lower())
    stop_words = set(stopwords.words('english'))
    palabras_filtradas = [p for p in palabras
                          if p.isalnum() and p not in stop_words]

    # Calcular frecuencia
    frecuencia = {}
    for palabra in palabras_filtradas:

```

```

frecuencia[palabra] = frecuencia.get(palabra, 0) + 1

# Normalizar frecuencias
max_freq = max(frecuencia.values()) if frecuencia else 1
frecuencia_norm = {k: v/max_freq for k, v in frecuencia.items()}

# Puntuar oraciones
puntajes = []
for oracion in oraciones:
    palabras_oracion = word_tokenize(oracion.lower())
    puntaje = sum(frecuencia_norm.get(p, 0)
                  for p in palabras_oracion
                  if p.isalnum())
    # Normalizar por longitud de oracion
    puntajes.append(puntaje / len(palabras_oracion)
                   if palabras_oracion else 0)

# Seleccionar top N oraciones
indices_top = sorted(range(len(puntajes)),
                     key=lambda i: puntajes[i],
                     reverse=True)[:num_oraciones]

# CLAVE: Reordenar cronologicamente
indices_top = sorted(indices_top)

resumen = [oraciones[i] for i in indices_top]
return resumen

```

Listing 9: Funciones de resumen automático (Parte 2: RAKE y TextRank)

```

def resumen_rake(texto, num_oraciones=12):
    """
    --- Genera resumen usando RAKE

    --- Logica: Extrae frases clave usando stopwords y puntuacion
    --- como delimitadores. Las oraciones con mas frases clave
    --- son mas importantes.

```



```

"""
    oraciones = sent_tokenize(texto)
    if len(oraciones) < num_oraciones:
        num_oraciones = len(oraciones)

    # Extraer frases clave con RAKE
    rake = Rake()
    rake.extract_keywords_from_text(texto)
    frases_clave = set([frase.lower()
                        for frase in rake.get_ranked_phrases()])

    # Puntuar oraciones segun presencia de frases clave
    puntajes = []
    for oracion in oraciones:
        oracion_lower = oracion.lower()
        puntaje = sum(1 for frase in frases_clave
                      if frase in oracion_lower)
        puntajes.append(puntaje)

    # Seleccionar top N oraciones
    indices_top = sorted(range(len(puntajes)),
                        key=lambda i: puntajes[i],
                        reverse=True)[:num_oraciones]

    # CLAVE: Reordenar cronologicamente
    indices_top = sorted(indices_top)

    resumen = [oraciones[i] for i in indices_top]
    return resumen

def resumen_textrank(texto, num_oraciones=12):
    """
    --- Genera resumen usando TextRank

    --- Logica: Construye un grafo donde los nodos son oraciones

```

```

----y-las-aristas-son-similitud-TF-IDF.-Aplica-PageRank-para
----encontrar-oraciones-centrales.
----"""

    oraciones = sent_tokenize(texto)
    if len(oraciones) < num_oraciones:
        num_oraciones = len(oraciones)

    # Vectorizar con TF-IDF
    tfidf_vectorizer = TfidfVectorizer()
    tfidf_matrix = tfidf_vectorizer.fit_transform(oraciones)

    # Calcular matriz de similitud
    similitud = cosine_similarity(tfidf_matrix, tfidf_matrix)

    # Crear grafo
    grafo = nx.from_numpy_array(similitud)

    # Aplicar PageRank
    puntajes = nx.pagerank(grafo)

    # Seleccionar top N oraciones
    indices_top = sorted(puntajes.keys(),
                        key=lambda i: puntajes[i],
                        reverse=True)[:num_oraciones]

    # CLAVE: Reordenar cronologicamente
    indices_top = sorted(indices_top)

    resumen = [oraciones[i] for i in indices_top]
    return resumen

```

Listing 10: Funciones de resumen automático (Parte 3: LSA y BERT)

```

def resumen_lsa(texto, num_oraciones=12):
    """
    ----Genera-resumen-usando-LSA-(Latent-Semantic-Analysis)

```

```

----Logica: Usa SVD para descubrir topicos latentes y puntua
----oraciones segun su relevancia para estos topicos.
----"""

    oraciones = sent_tokenize(texto)
    if len(oraciones) < num_oraciones:
        num_oraciones = len(oraciones)

    # Vectorizar con CountVectorizer
    vectorizer = CountVectorizer(stop_words='english')
    document_matrix = vectorizer.fit_transform(oraciones)

    # Aplicar SVD
    n_components = min(num_oraciones, len(oraciones),
                        document_matrix.shape[1])
    svd = TruncatedSVD(n_components=n_components)
    document_matrix_reduced = svd.fit_transform(document_matrix)

    # Puntuar oraciones (suma de valores absolutos en espacio latente)
    puntajes = [sum(abs(document_matrix_reduced[i]))
                 for i in range(len(oraciones))]

    # Seleccionar top N oraciones
    indices_top = sorted(range(len(puntajes)),
                          key=lambda i: puntajes[i],
                          reverse=True)[:num_oraciones]

    # CLAVE: Reordenar cronologicamente
    indices_top = sorted(indices_top)

    resumen = [oraciones[i] for i in indices_top]
    return resumen

def resumen_bert(texto, num_oraciones=12, modelo=None):
    """
    ----Genera resumen usando BERT (Transformers)

```

```

---Logica: -Genera embeddings para el documento completo y
---para cada oracion. Las oraciones mas similares al documento
---(similitud coseno) son las mas representativas.

---OPTIMIZACION: Recibe el modelo pre-cargado como parametro
---"""
    if modelo is None:
        modelo = MODELO.BERT

    oraciones = sent_tokenize(texto)
    if len(oraciones) < num_oraciones:
        num_oraciones = len(oraciones)

    # Generar embedding del documento completo
    doc_embedding = modelo.encode([texto])[0]

    # Generar embeddings de cada oracion
    oraciones_embeddings = modelo.encode(oraciones)

    # Calcular similitud coseno con el documento
    puntajes = [
        cosine_similarity([doc_embedding], [sent_emb])[0][0]
        for sent_emb in oraciones_embeddings
    ]

    # Seleccionar top N oraciones
    indices_top = sorted(range(len(puntajes)),
                        key=lambda i: puntajes[i],
                        reverse=True)[:num_oraciones]

    # CLAVE: Reordenar cronologicamente
    indices_top = sorted(indices_top)

    resumen = [oraciones[i] for i in indices_top]
    return resumen

```

```
print(" - Funciones de resumen definidas (12 oraciones con optimizaciones) ")
```

Explicación de cada algoritmo:

1. TF-IDF:

- Vectoriza cada oración usando `TfidfVectorizer`
- Suma los valores TF-IDF de todas las palabras en cada oración
- Las oraciones con palabras distintivas (altos valores TF-IDF) obtienen mayor puntaje

2. Frecuencia:

- Calcula frecuencia de cada palabra (excluyendo stopwords)
- Normaliza dividiendo por la frecuencia máxima
- Puntúa cada oración como el promedio de frecuencias normalizadas de sus palabras
- Normaliza por longitud para evitar sesgo hacia oraciones largas

3. RAKE:

- Extrae frases clave automáticamente usando delimitadores naturales
- Cuenta cuántas frases clave aparecen en cada oración
- Las oraciones con más frases clave son más importantes

4. TextRank:

- Crea un grafo donde nodos = oraciones, aristas = similitud TF-IDF
- Aplica algoritmo PageRank (mismo que usa Google)
- Las oraciones centrales (conectadas a muchas otras importantes) ganan

5. LSA:

- Crea matriz término-documento y aplica SVD
- Descubre tópicos latentes (dimensiones ocultas del significado)
- Puntúa oraciones según su presencia en el espacio de tópicos

6. BERT:

- Genera embedding del documento completo (vector de 384 dimensiones)
- Genera embedding de cada oración individual
- Calcula similitud coseno entre cada oración y el documento
- **Optimización:** Recibe modelo pre-cargado para evitar redundancia

4.3.3. Ejecución y Medición de Tiempos

Se ejecutan los seis algoritmos para cada carta y se mide el tiempo de ejecución:

Listing 11: Ejecución de algoritmos y medición de tiempos

```
# =====
# EJECUTAR EXTRACCION Y NORMALIZACION FINAL
# =====

print("\n" + "="*80)
print("EXTRACCION-DE-CARTAS-DE-FRANKENSTEIN")
print("="*80 + "\n")

cartas = obtener_cartas_frankenstein()

print("\n" + "="*80)
print(f"RESULTADO-FINAL: {len(cartas)}-cartas-extraidas")
print("="*80 + "\n")

# Normalizar para cada algoritmo
cartas_tfidf = [normalizar_para_tfidf(carta) for carta in cartas]
cartas_frecuencia = [normalizar_para_frecuencia(carta) for carta in cartas]
cartas_rake = [normalizar_para_rake(carta) for carta in cartas]
cartas_textrank = [normalizar_para_textrank(carta) for carta in cartas]
cartas_bert = [normalizar_para_bert(carta) for carta in cartas]
cartas_lsa = [normalizar_para_lsa(carta) for carta in cartas]

print("- Textos-normalizados-para-cada-algoritmo\n")

# =====
```

```
# EJECUCION DE ALGORITMOS Y MEDICION DE TIEMPOS (PUNTO 3)
# =====
print("="*80)
print("EJECUTANDO ALGORITMOS DE RESUMEN Y MIDIENDO TIEMPOS")
print("="*80 + "\n")

# Inicializar almacenamiento de tiempos y resúmenes
tiempos_tfidf = []
tiempos_frecuencia = []
tiempos_rake = []
tiempos_textrank = []
tiempos_bert = []
tiempos_lsa = []

resúmenes_tfidf = []
resúmenes_frecuencia = []
resúmenes_rake = []
resúmenes_textrank = []
resúmenes_bert = []
resúmenes_lsa = []

# Ejecutar para cada carta
for i in range(len(cartas)):
    print(f"Procesando Carta-{i+1}...")

    # TF-IDF
    inicio = time.time()
    res = resumen_tfidf(cartas_tfidf[i])
    tiempo = time.time() - inicio
    tiempos_tfidf.append(tiempo)
    resúmenes_tfidf.append(res)
    print(f"TF-IDF: {tiempo:.4f}s")

    # Frecuencia
    inicio = time.time()
    res = resumen_frecuencia(cartas_frecuencia[i])
```

```
tiempo = time.time() - inicio
tiempos_frecuencia.append(tiempo)
resumenes_frecuencia.append(res)
print(f" - Frecuencia: -{tiempo:.4f}s")

# RAKE
inicio = time.time()
res = resumen_rake(cartas_rake[i])
tiempo = time.time() - inicio
tiempos_rake.append(tiempo)
resumenes_rake.append(res)
print(f" - RAKE: -{tiempo:.4f}s")

# TextRank
inicio = time.time()
res = resumen_textrank(cartas_textrank[i])
tiempo = time.time() - inicio
tiempos_textrank.append(tiempo)
resumenes_textrank.append(res)
print(f" - TextRank: -{tiempo:.4f}s")

# BERT (usando modelo pre-cargado)
inicio = time.time()
res = resumen_bert(cartas_bert[i], modelo=MODELO.BERT)
tiempo = time.time() - inicio
tiempos_bert.append(tiempo)
resumenes_bert.append(res)
print(f" - BERT: -{tiempo:.4f}s")

# LSA
inicio = time.time()
res = resumen_lsa(cartas_lsa[i])
tiempo = time.time() - inicio
tiempos_lsa.append(tiempo)
resumenes_lsa.append(res)
print(f" - LSA: -{tiempo:.4f}s")
```



```
print()  
  
print("="*80)  
print("-EJECUCION-COMPLETADA")  
print("="*80)
```

Resultado de la ejecución:

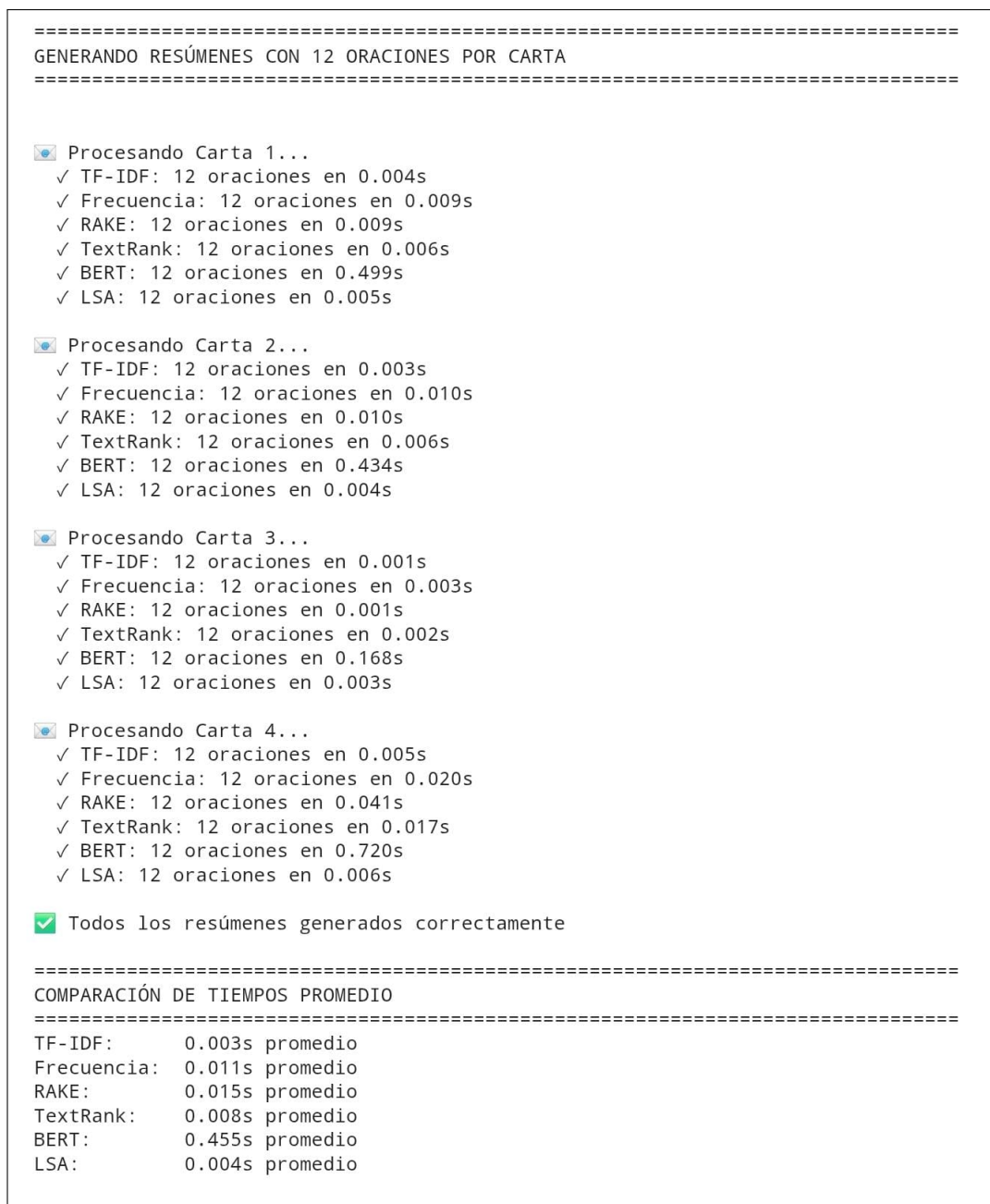


Figura 6: Tiempos de ejecución de los 6 algoritmos para las 4 cartas

Interpretación preliminar: Se observa que BERT es significativamente más lento (~ 0.5 segundos por carta) en comparación con los métodos estadísticos como TF-IDF (~ 0.003 segundos). Esta diferencia se analizará en detalle en el Punto 4.

4.3.4. Visualización de Resúmenes Generados

Para validar la calidad de los resúmenes, se muestran las 12 oraciones seleccionadas por cada algoritmo para cada carta:

Listing 12: Visualización de resúmenes generados

```
# =====
# VISUALIZACION DE RESUMENES GENERADOS
# =====

algoritmos = [ 'TF-IDF', 'Frecuencia', 'RAKE', 'TextRank', 'BERT', 'LSA' ]
todos_resumenes = [ resúmenes_tfidf, resúmenes_frecuencia, resúmenes_rake,
                    resúmenes_textrank, resúmenes_bert, resúmenes_lsa ]

for i in range(len(cartas)):
    print("\n" + "="*80)
    print(f"RESUMENES-DE-CARTA-{i+1}")
    print("="*80 + "\n")

    for j, algoritmo in enumerate(algoritmos):
        print(f"\n{' '*80}")
        print(f"ALGORITMO: {algoritmo}")
        print(f"{' '*80}\n")

        resumen = todos_resumenes[j][i]
        for k, oracion in enumerate(resumen, 1):
            print(f"{k}. {oracion}")
            print()

        print(f"{' '*80}")
        print(f"Total: {len(resumen)} oraciones")
        print(f"{' '*80}\n")

    print("\n" + "="*80)
    print(f"VISUALIZACION-COMPLETADA--RESUMENES-GENERADOS")
    print("="*80)
```

Resúmenes generados por los 6 algoritmos:



=====
RESÚMENES GENERADOS - 12 ORACIONES POR CARTA
=====
CARTA 1 - RESÚMENES DE TODOS LOS ALGORITMOS
=====
 TF-IDF
=====
1. I am already far north of London, and as I walk in the streets of Petersburg, I feel a cold northern breeze play upon my cheeks, which braces my nerves and fills me with delight.
2. There for with your leave, my sister, I will put some trust in preceding navigators there snow and frost are banished and, sailing over a calm sea, we may be wafted to a land surpassing in wonders and in beauty every region hitherto discovered on the habitable globe.
3. I may there discover the wondrous power which attracts the needle and may regulate a thousand celestial observations that require only this voyage to render their seeming eccentricities consistent for ever.
4. These are my enticements, and they are sufficient to conquer all fear of danger or death and to induce me to commence this laborious voyage with the joy a child feels when he embarks in a little boat, with his holiday mates, on an expedition of discovery up his native river.
5. But supposing all these conjectures to be false, you cannot contest the inestimable benefit which I shall confer on all mankind, to the last generation, by discovering a passage near the pole to those countries, to reach which at present so many months are requisite or by ascertaining the secret of the magnet, which, if at all possible, can only be effected by an undertaking such as mine.
6. These reflections have dispelled the agitation with which I began my letter, and I feel my heart glow with an enthusiasm which elevates me to heaven, for nothing contributes so much to tranquillise the mind as a steady purpose a point on which the soul may fix its intellectual eye.
7. These volumes were my study day and night, and my familiarity with them increased that regret which I had felt, as a child, on learning that my father's dying injunction had forbidden my uncle to allow me to embark in a seafaring life.
8. I accompanied the whale fishers on several expeditions to the North Sea I voluntarily endured cold, famine, thirst, and want of sleep I often worked harder than the common sailors during the day and devoted my nights to the study of mathematics, the theory of medicine, and those branches of physical science from which a naval adventurer might derive the greatest practical advantage.
9. I must own I felt a little proud when my captain offered me the second dignity in the vessel and entreated me to remain with the greatest earnestness, so valuable did he consider my services.
10. I am about to proceed on a long and difficult voyage, the emergencies of which will demand all my fortitude I am required not only to raise the spirits of others, but sometimes to sustain my own, when theirs are failing.
11. The cold is not excessive, if you are wrapped in furs a dress which I have already adopted, for there is a great difference between walking the deck and remaining seated motionless for hours, when no exercise prevents the blood from actually freezing in your veins.
12. I shall depart for the latter town in a fortnight or three weeks and my intention is to hire a ship there, which can easily be done by paying the insurance for the owner, and to engage as many sailors as I think necessary among those who are accustomed to the whale fishing.
Total: 12 oraciones
=====
 Frecuencia
=====
1. St. Petersburg, Dec. 11th, 17 .
2. Its productions and features may be without example, as the phenomena of the heavenly bodies undoubtedly are in those undiscovered solitudes.
3. What may not be expected in a country of eternal light?
4. I may there discover the wondrous power which attracts the needle and may regulate a thousand celestial observations that require only this voyage to render their seeming eccentricities consistent for ever.
5. I shall satiate my ardent curiosity with the sight of a part of the world never before visited, and may tread a land never before imprinted by the foot of man.
6. You may remember that a history of all the voyages made for purposes of discovery composed the whole of our good Uncle Thomas library.
7. Six years have passed since I resolved on my present undertaking.
8. And now, dear Margaret, do I not deserve to accomplish some great purpose?
9. Ah, dear sister, how can I answer this question?
10. If I succeed, many, many months, perhaps years, will pass before you and I may meet.
11. Farewell, my dear, excellent Margaret.
12. Heaven shower down blessings on you, and save me, that I may again and again testify my gratitude for all your love and kindness.
Total: 12 oraciones
=====

Figura 7: Resúmenes de Carta 1 - Algoritmos TF-IDF y Frecuencia



<p>=====</p> <p> RAKE</p> <p>=====</p> <ol style="list-style-type: none"> 1. I am already far north of London, and as I walk in the streets of Petersburg, I feel a cold northern breeze play upon my cheeks, which braces my nerves and fills me with delight. 2. There-for with your leave, my sister, I will put some trust in preceding navigators—there s now and frost are banished; and, sailing over a calm sea, we may be wafted to a land surpassin g in wonders and in beauty every region hitherto discovered on the habitable globe. 3. I may there discover the wondrous power which attracts the needle and may regulate a thousa nd celestial observations that require only this voyage to render their seeming eccentricities consistent for ever. 4. I shall satiate my ardent curiosity with the sight of a part of the world never before visi ted, and may tread a land never before imprinted by the foot of man. 5. These are my enticements, and they are sufficient to conquer all fear of danger or death an d to induce me to commence this laborious voyage with the joy a child feels when he embarks in a little boat, with his holiday mates, on an expedition of discovery up his native river. 6. But supposing all these conjectures to be false, you cannot contest the inestimable benefit which I shall confer on all mankind, to the last generation, by discovering a passage near the pole to those countries, to reach which at present so many months are requisite; or by ascerta ining the secret of the magnet, which, if at all possible, can only be effected by an undertak ing such as mine. 7. These reflections have dispelled the agitation with which I began my letter, and I feel my heart glow with an enthusiasm which elevates me to heaven, for nothing contributes so much to tranquillise the mind as a steady purpose—a point on which the soul may fix its intellectual e ye. 8. These volumes were my study day and night, and my familiarity with them increased that regr et which I had felt, as a child, on learning that my father's dying injunction had forbidden m y uncle to allow me to embark in a seafaring life. 9. These visions faded when I perused, for the first time, those poets whose effusions entranc ed my soul and lifted it to heaven. 10. I accompanied the whale-fishers on several expeditions to the North Sea; I voluntarily end ured cold, famine, thirst, and want of sleep; I often worked harder than the common sailors du ring the day and devoted my nights to the study of mathematics, the theory of medicine, and th ose branches of physical science from which a naval adventurer might derive the greatest pract ical advantage. 11. The cold is not excessive, if you are wrapped in furs—a dress which I have already adopte d, for there is a great difference between walking the deck and remaining seated motionless fo r hours, when no exercise prevents the blood from actually freezing in your veins. 12. I shall depart for the latter town in a fortnight or three weeks; and my intention is to h ire a ship there, which can easily be done by paying the insurance for the owner, and to engag e as many sailors as I think necessary among those who are accustomed to the whale-fishing. <p>Total: 12 oraciones</p> <p>=====</p>	<p>=====</p> <p> TextRank</p> <p>=====</p> <ol style="list-style-type: none"> 1. I arrived here yesterday, and my first task is to assure my dear sister of my welfare and i ncreasing confidence in the success of my undertaking. 2. I am already far north of London, and as I walk in the streets of Petersburg, I feel a col d northern breeze play upon my cheeks, which braces my nerves and fills me with delight. 3. I try in vain to be persuaded that the pole is the seat of frost and desolation it ever pre sents itself to my imagination as the region of beauty and delight. 4. There for with your leave, my sister, I will put some trust in preceding navigators there s now and frost are banished and, sailing over a calm sea, we may be wafted to a land surpassing in wonders and in beauty every region hitherto discovered on the habitable globe. 5. These are my enticements, and they are sufficient to conquer all fear of danger or death an d to induce me to commence this laborious voyage with the joy a child feels when he embarks in a little boat, with his holiday mates, on an expedition of discovery up his native river. 6. But supposing all these conjectures to be false, you cannot contest the inestimable benefit which I shall confer on all mankind, to the last generation, by discovering a passage near the pole to those countries, to reach which at present so many months are requisite or by ascertai ning the secret of the magnet, which, if at all possible, can only be effected by an undertaki ng such as mine. 7. These reflections have dispelled the agitation with which I began my letter, and I feel my heart glow with an enthusiasm which elevates me to heaven, for nothing contributes so much to tranquillise the mind as a steady purpose a point on which the soul may fix its intellectual e ye. 8. I have read with ardour the accounts of the various voyages which have been made in the pro spect of arriving at the North Pacific Ocean through the seas which surround the pole. 9. I accompanied the whale fishers on several expeditions to the North Sea I voluntarily endur ed cold, famine, thirst, and want of sleep I often worked harder than the common sailors durin g the day and devoted my nights to the study of mathematics, the theory of medicine, and those branches of physical science from which a naval adventurer might derive the greatest practical advantage. 10. I am about to proceed on a long and difficult voyage, the emergencies of which will demand all my fortitude I am required not only to raise the spirits of others, but sometimes to susta in my own, when theirs are failing. 11. The cold is not excessive, if you are wrapped in furs a dress which I have already adopte d, for there is a great difference between walking the deck and remaining seated motionless fo r hours, when no exercise prevents the blood from actually freezing in your veins. 12. I shall depart for the latter town in a fortnight or three weeks and my intention is to hi re a ship there, which can easily be done by paying the insurance for the owner, and to engag e as many sailors as I think necessary among those who are accustomed to the whale fishing. <p>Total: 12 oraciones</p> <p>=====</p>
---	---

Figura 8: Resúmenes de Carta 1 - Algoritmos RAKE y TextRank



<p>=====</p> <p> BERT</p> <p>=====</p> <p>1. _To Mrs. Saville, England._ St. Petersburg, Dec. 11th, 17-.</p> <p>2. I arrived here yesterday, and my first task is to assure my dear sister of my welfare and increasing confidence in the success of my undertaking.</p> <p>3. I am already far north of London, and as I walk in the streets of Petersburg, I feel a cold northern breeze play upon my cheeks, which braces my nerves and fills me with delight.</p> <p>4. This breeze, which has travelled from the regions towards which I am advancing, gives me a foretaste of those icy climes.</p> <p>5. Inspired by this wind of promise, my daydreams become more fervent and vivid.</p> <p>6. I try in vain to be persuaded that the pole is the seat of frost and desolation; it ever presents itself to my imagination as the region of beauty and delight.</p> <p>7. There, Margaret, the sun is for ever visible, its broad disk just skirting the horizon and diffusing a perpetual splendour.</p> <p>8. There-for with your leave, my sister, I will put some trust in preceding navigators-there snow and frost are banished; and, sailing over a calm sea, we may be wafted to a land surpassing in wonders and in beauty every region hitherto discovered on the habitable globe.</p> <p>9. I shall satiate my ardent curiosity with the sight of a part of the world never before visited, and may tread a land never before imprinted by the foot of man.</p> <p>10. These are my enticements, and they are sufficient to conquer all fear of danger or death and to induce me to commence this laborious voyage with the joy a child feels when he embarks in a little boat, with his holiday mates, on an expedition of discovery up his native river.</p> <p>11. These reflections have dispelled the agitation with which I began my letter, and I feel my heart glow with an enthusiasm which elevates me to heaven, for nothing contributes so much to tranquillise the mind as a steady purpose-a point on which the soul may fix its intellectual eye.</p> <p>12. I am about to proceed on a long and difficult voyage, the emergencies of which will demand all my fortitude: I am required not only to raise the spirits of others, but sometimes to sustain my own, when theirs are failing.</p> <p>Total: 12 oraciones</p> <p>-----</p>	<p>=====</p> <p> LSA</p> <p>=====</p> <p>1. i am already far north of london, and as i walk in the streets of petersburgh, i feel a cold northern breeze play upon my cheeks, which braces my nerves and fills me with delight.</p> <p>2. there for with your leave, my sister, i will put some trust in preceding navigators there snow and frost are banished and, sailing over a calm sea, we may be wafted to a land surpassing in wonders and in beauty every region hitherto discovered on the habitable globe.</p> <p>3. i may there discover the wondrous power which attracts the needle and may regulate a thousand celestial observations that require only this voyage to render their seeming eccentricities consistent for ever.</p> <p>4. these are my enticements, and they are sufficient to conquer all fear of danger or death and to induce me to commence this laborious voyage with the joy a child feels when he embarks in a little boat, with his holiday mates, on an expedition of discovery up his native river.</p> <p>5. but supposing all these conjectures to be false, you cannot contest the inestimable benefit which i shall confer on all mankind, to the last generation, by discovering a passage near the pole to those countries, to reach which at present so many months are requisite or by ascertaining the secret of the magnet, which, if at all possible, can only be effected by an undertaking such as mine.</p> <p>6. these reflections have dispelled the agitation with which i began my letter, and i feel my heart glow with an enthusiasm which elevates me to heaven, for nothing contributes so much to tranquillise the mind as a steady purpose a point on which the soul may fix its intellectual eye.</p> <p>7. i have read with ardour the accounts of the various voyages which have been made in the prospect of arriving at the north pacific ocean through the seas which surround the pole.</p> <p>8. these volumes were my study day and night, and my familiarity with them increased that regret which i had felt, as a child, on learning that my father's dying injunction had forbidden my uncle to allow me to embark in a seafaring life.</p> <p>9. i accompanied the whale fishers on several expeditions to the north sea i voluntarily endured cold, famine, thirst, and want of sleep i often worked harder than the common sailors during the day and devoted my nights to the study of mathematics, the theory of medicine, and those branches of physical science from which a naval adventurer might derive the greatest practical advantage.</p> <p>10. i must own i felt a little proud when my captain offered me the second dignity in the vessel and entreated me to remain with the greatest earnestness, so valuable did he consider my services.</p> <p>11. the cold is not excessive, if you are wrapped in furs a dress which i have already adopted, for there is a great difference between walking the deck and remaining seated motionless for hours, when no exercise prevents the blood from actually freezing in your veins.</p> <p>12. i shall depart for the latter town in a fortnight or three weeks and my intention is to hire a ship there, which can easily be done by paying the insurance for the owner, and to engage as many sailors as i think necessary among those who are accustomed to the whale fishing.</p> <p>Total: 12 oraciones</p> <p>-----</p>
---	--

Figura 9: Resúmenes de Carta 1 - Algoritmos BERT y LSA



<p>-----</p> <p>CARTA 2 - RESÚMENES DE TODOS LOS ALGORITMOS</p> <p>=====</p> <p> TF-IDF</p> <p>=====</p> <p>1. But I have one want which I have never yet been able to satisfy, and the absence of the object of which I now feel as a most severe evil, I have no friend, Margaret when I am glowing with the enthusiasm of success, there will be none to participate my joy if I am assailed by disappointment, no one will endeavour to sustain me in dejection.</p> <p>2. I have no one near me, gentle yet courageous, possessed of a cultivated as well as of a capacious mind, whose tastes are like my own, to approve or amend my plans.</p> <p>3. But it is a still greater evil to me that I am self educated for the first fourteen years of my life I ran wild on a common and read nothing but our Uncle Thomas books of voyages.</p> <p>4. At that age I became acquainted with the celebrated poets of our own country but it was only when it had ceased to be in my power to derive its most important benefits from such a conviction that I perceived the necessity of becoming acquainted with more languages than that of my native country.</p> <p>5. It is true that I have thought more and that my daydreams are more extended and magnificent, but they want as the painters call it and I greatly need a friend who would have sense enough not to despise me as romantic, and affection enough for me to endeavour to regulate my mind.</p> <p>6. A youth passed in solitude, my best years spent under your gentle and feminine fosterage, has so refined the groundwork of my character that I cannot overcome an intense distaste to the usual brutality exercised on board ship I have never believed it to be necessary, and when I heard of a mariner equally noted for his kindness of heart and the respect and obedience paid to him by his crew, I felt myself peculiarly fortunate in being able to secure his services.</p> <p>7. He saw his mistress once before the destined ceremony but she was bathed in tears, and throwing herself at his feet, entreated him to spare her, confessing at the same time that she loved another, but that he was poor, and that her father would never consent to the union.</p> <p>8. He had already bought a farm with his money, on which he had designed to pass the remainder of his life but he bestowed the whole on his rival, together with the remains of his prize money to purchase stock, and then himself solicited the young woman's father to consent to her marriage with her lover.</p> <p>9. But the old man decidedly refused, thinking himself bound in honour to my friend, who, when he found the father inexorable, quitted his country, nor returned until he heard that his former mistress was married according to her inclinations.</p> <p>10. He is so but then he is wholly uneducated he is as silent as a Turk, and a kind of ignorant carelessness attends him, which, while it renders his conduct the more astonishing, detracts from the interest and sympathy which otherwise he would command.</p> <p>11. I am going to unexplored regions, to the land of mist and snow, but I shall kill no albatross therefore do not be alarmed for my safety or if I should come back to you as worn and woeful as the Ancient Mariner.</p> <p>12. I am practically industrious painstaking, a workman to execute with perseverance and labour but besides this there is a love for the marvellous, a belief in the marvellous, intertwined in all my projects, which hurries me out of the common pathways of men, even to the wild sea and unvisited regions I am about to explore.</p> <p>Total: 12 oraciones</p> <p>-----</p> <p> Frecuencia</p> <p>=====</p> <p>1. Yet a second step is taken towards my enterprise.</p> <p>2. But I have one want which I have never yet been able to satisfy, and the absence of the object of which I now feel as a most severe evil, I have no friend, Margaret when I am glowing with the enthusiasm of success, there will be none to participate my joy if I am assailed by disappointment, no one will endeavour to sustain me in dejection.</p> <p>3. I shall commit my thoughts to paper, it is true but that is a poor medium for the communication of feeling.</p> <p>4. You may deem me romantic, my dear sister, but I bitterly feel the want of a friend.</p> <p>5. I have no one near me, gentle yet courageous, possessed of a cultivated as well as of a capacious mind, whose tastes are like my own, to approve or amend my plans.</p> <p>6. How would such a friend repair the faults of your poor brother!</p> <p>7. It is true that I have thought more and that my daydreams are more extended and magnificent, but they want as the painters call it and I greatly need a friend who would have sense enough not to despise me as romantic, and affection enough for me to endeavour to regulate my mind.</p> <p>8. Well, these are useless complaints I shall certainly find no friend on the wide ocean, nor even here in Archangel, among merchants and seamen.</p> <p>9. I heard of him first in rather a romantic manner, from a lady who owes to him the happiness of her life.</p> <p>10. But the old man decidedly refused, thinking himself bound in honour to my friend, who, when he found the father inexorable, quitted his country, nor returned until he heard that his former mistress was married according to her inclinations.</p> <p>11. Yet do not suppose, because I complain a little or because I can conceive a consolation for my toils which I may never know, that I am wavering in my resolutions.</p> <p>12. Remember me with affection, should you never hear from me again.</p> <p>Total: 12 oraciones</p> <p>-----</p>	
---	--

Figura 10: Resúmenes de Carta 2 - Algoritmos TF-IDF y Frecuencia



<p>=====</p> <p> RAKE</p> <p>=====</p> <p>1. But I have one want which I have never yet been able to satisfy, and the absence of the object of which I now feel as a most severe evil, I have no friend, Margaret: when I am glowing with the enthusiasm of success, there will be none to participate my joy; if I am assailed by disappointment, no one will endeavour to sustain me in dejection.</p> <p>2. I have no one near me, gentle yet courageous, possessed of a cultivated as well as of a capacious mind, whose tastes are like my own, to approve or amend my plans.</p> <p>3. It is true that I have thought more and that my daydreams are more extended and magnificent, but they want (as the painters call it) and I greatly need a friend who would have sense enough not to despise me as romantic, and affection enough for me to endeavour to regulate my mind.</p> <p>4. Well, these are useless complaints: I shall certainly find no friend on the wide ocean, nor even here in Archangel, among merchants and seamen.</p> <p>5. A youth passed in solitude, my best years spent under your gentle and feminine fosterage, has so refined the groundwork of my character that I cannot overcome an intense distaste to the usual brutality exercised on board ship: I have never believed it to be necessary, and when I heard of a mariner equally noted for his kindness of heart and the respect and obedience paid to him by his crew, I felt myself peculiarly fortunate in being able to secure his services.</p> <p>6. Some years ago he loved a young Russian lady of moderate fortune, and having amassed a considerable sum in prize-money, the father of the girl consented to the match.</p> <p>7. He saw his mistress once before the destined ceremony; but she was bathed in tears, and throwing herself at his feet, entreated him to spare her, confessing at the same time that she loved another, but that he was poor, and that her father would never consent to the union.</p> <p>8. He had already bought a farm with his money, on which he had designed to pass the remainder of his life; but he bestowed the whole on his rival, together with the remains of his prize-money to purchase stock, and then himself solicited the young woman's father to consent to her marriage with her lover.</p> <p>9. But the old man decidedly refused, thinking himself bound in honour to my friend, who, when he found the father inexorable, quitted his country, nor returned until he heard that his former mistress was married according to her inclinations.</p> <p>10. He is so; but then he is wholly uneducated: he is as silent as a Turk, and a kind of ignorant carelessness attends him, which, while it renders his conduct the more astonishing, detracts from the interest and sympathy which otherwise he would command.</p> <p>11. The winter has been dreadfully severe, but the spring promises well, and it is considered as a remarkably early season, so that perhaps I may sail sooner than I expected.</p> <p>12. I am going to unexplored regions, to "the land of mist and snow," but I shall kill no albatross; therefore do not be alarmed for my safety or if I should come back to you as worn and woeful as the "Ancient Mariner." You will smile at my allusion, but I will disclose a secret.</p> <p>Total: 12 oraciones</p> <p>-----</p>	<p>=====</p> <p> TextRank</p> <p>=====</p> <p>1. But I have one want which I have never yet been able to satisfy, and the absence of the object of which I now feel as a most severe evil, I have no friend, Margaret when I am glowing with the enthusiasm of success, there will be none to participate my joy if I am assailed by disappointment, no one will endeavour to sustain me in dejection.</p> <p>2. At that age I became acquainted with the celebrated poets of our own country but it was only when it had ceased to be in my power to derive its most important benefits from such a conviction that I perceived the necessity of becoming acquainted with more languages than that of my native country.</p> <p>3. It is true that I have thought more and that my daydreams are more extended and magnificent, but they want as the painters call it and I greatly need a friend who would have sense enough not to despise me as romantic, and affection enough for me to endeavour to regulate my mind.</p> <p>4. My lieutenant, for instance, is a man of wonderful courage and enterprise he is madly desirous of glory, or rather, to word my phrase more characteristically, of advancement in his profession.</p> <p>5. The master is a person of an excellent disposition and is remarkable in the ship for his gentleness and the mildness of his discipline.</p> <p>6. A youth passed in solitude, my best years spent under your gentle and feminine fosterage, has so refined the groundwork of my character that I cannot overcome an intense distaste to the usual brutality exercised on board ship I have never believed it to be necessary, and when I heard of a mariner equally noted for his kindness of heart and the respect and obedience paid to him by his crew, I felt myself peculiarly fortunate in being able to secure his services.</p> <p>7. He saw his mistress once before the destined ceremony but she was bathed in tears, and throwing herself at his feet, entreated him to spare her, confessing at the same time that she loved another, but that he was poor, and that her father would never consent to the union.</p> <p>8. He had already bought a farm with his money, on which he had designed to pass the remainder of his life but he bestowed the whole on his rival, together with the remains of his prize money to purchase stock, and then himself solicited the young woman's father to consent to her marriage with her lover.</p> <p>9. I shall do nothing rashly you know me sufficiently to confide in my prudence and consideration whenever the safety of others is committed to my care.</p> <p>10. I am going to unexplored regions, to the land of mist and snow, but I shall kill no albatross therefore do not be alarmed for my safety or if I should come back to you as worn and woeful as the Ancient Mariner.</p> <p>11. I have often attributed my attachment to, my passionate enthusiasm for, the dangerous mysteries of ocean to that production of the most imaginative of modern poets.</p> <p>12. I am practically industrious painstaking, a workman to execute with perseverance and labour but besides this there is a love for the marvellous, a belief in the marvellous, intertwined in all my projects, which hurries me out of the common pathways of men, even to the wild sea and unvisited regions I am about to explore.</p> <p>Total: 12 oraciones</p> <p>-----</p>
--	--

Figura 11: Resúmenes de Carta 2 - Algoritmos RAKE y TextRank



<p>=====</p> <p> BERT</p> <p>=====</p> <p>1. _To Mrs. Saville, England._ Archangel, 28th March, 17-.</p> <p>2. I have hired a vessel and am occupied in collecting my sailors: those whom I have already engaged appear to be men on whom I can depend and are certainly possessed of dauntless courage.</p> <p>3. But I have one want which I have never yet been able to satisfy, and the absence of the object of which I now feel as a most severe evil, I have no friend, Margaret: when I am glowing with the enthusiasm of success, there will be none to participate my joy; if I am assailed by disappointment, no one will endeavour to sustain me in dejection.</p> <p>4. You may deem me romantic, my dear sister, but I bitterly feel the want of a friend.</p> <p>5. I have no one near me, gentle yet courageous, possessed of a cultivated as well as of a capacious mind, whose tastes are like my own, to approve or amend my plans.</p> <p>6. It is true that I have thought more and that my daydreams are more extended and magnificent, but they want (as the painters call it) keeping; and I greatly need a friend who would have sense enough not to despise me as romantic, and affection enough for me to endeavour to regulate my mind.</p> <p>7. Well, these are useless complaints; I shall certainly find no friend on the wide ocean, nor even here in Archangel, among merchants and seamen.</p> <p>8. A youth passed in solitude, my best years spent under your gentle and feminine fosterage, has so refined the groundwork of my character that I cannot overcome an intense distaste to the usual brutality exercised on board ship: I have never believed it to be necessary, and when I heard of a mariner equally noted for his kindness of heart and the respect and obedience paid to him by his crew, I felt myself peculiarly fortunate in being able to secure his services.</p> <p>9. My generous friend reassured the suppliant, and on being informed of the name of her lover, instantly abandoned his pursuit.</p> <p>10. I shall do nothing rashly: you know me sufficiently to confide in my prudence and considerateness whenever the safety of others is committed to my care.</p> <p>11. I have often attributed my attachment to, my passionate enthusiasm for, the dangerous mysteries of ocean to that production of the most imaginative of modern poets.</p> <p>12. Continue for the present to write to me by every opportunity: I may receive your letters on some occasions when I need them most to support my spirits.</p> <p>Total: 12 oraciones</p> <p>-----</p>	<p>=====</p> <p> LSA</p> <p>=====</p> <p>1. but i have one want which i have never yet been able to satisfy, and the absence of the object of which i now feel as a most severe evil, i have no friend, margaret when i am glowing with the enthusiasm of success, there will be none to participate my joy if i am assailed by disappointment, no one will endeavour to sustain me in dejection.</p> <p>2. but it is a still greater evil to me that i am self educated for the first fourteen years of my life i ran wild on a common and read nothing but our uncle thomas books of voyages.</p> <p>3. at that age i became acquainted with the celebrated poets of our own country but it was only when it had ceased to be in my power to derive its most important benefits from such a conviction that i perceived the necessity of becoming acquainted with more languages than that of my native country.</p> <p>4. it is true that i have thought more and that my daydreams are more extended and magnificent, but they want as the painters call it and i greatly need a friend who would have sense enough not to despise me as romantic, and affection enough for me to endeavour to regulate my mind.</p> <p>5. my lieutenant, for instance, is a man of wonderful courage and enterprise he is madly desirous of glory, or rather, to word my phrase more characteristically, of advancement in his profession.</p> <p>6. a youth passed in solitude, my best years spent under your gentle and feminine fosterage, has so refined the groundwork of my character that i cannot overcome an intense distaste to the usual brutality exercised on board ship i have never believed it to be necessary, and when i heard of a mariner equally noted for his kindness of heart and the respect and obedience paid to him by his crew, i felt myself peculiarly fortunate in being able to secure his services.</p> <p>7. some years ago he loved a young russian lady of moderate fortune, and having amassed a considerable sum in prize money, the father of the girl consented to the match.</p> <p>8. he saw his mistress once before the destined ceremony but she was bathed in tears, and throwing herself at his feet, entreated him to spare her, confessing at the same time that she loved another, but that he was poor, and that her father would never consent to the union.</p> <p>9. he had already bought a farm with his money, on which he had designed to pass the remainder of his life but he bestowed the whole on his rival, together with the remains of his prize money to purchase stock, and then himself solicited the young woman's father to consent to her marriage with her lover.</p> <p>10. but the old man decidedly refused, thinking himself bound in honour to my friend, who, when he found the father inexorable, quitted his country, nor returned until he heard that his former mistress was married according to her inclinations.</p> <p>11. i am going to unexplored regions, to the land of mist and snow, but i shall kill no albatross therefore do not be alarmed for my safety or if i should come back to you as worn and woeful as the ancient mariner.</p> <p>12. i am practically industrious painstaking, a workman to execute with perseverance and labour but besides this there is a love for the marvellous, a belief in the marvellous, intertwined in all my projects, which hurries me out of the common pathways of men, even to the wild sea and unvisited regions i am about to explore.</p> <p>Total: 12 oraciones</p> <p>-----</p>
---	---

Figura 12: Resúmenes de Carta 2 - Algoritmos BERT y LSA



<p>-----</p> <p>CARTA 3 - RESÚMENES DE TODOS LOS ALGORITMOS</p> <p>=====</p> <p> TF-IDF</p> <p>=====</p> <p>1. My dear Sister, I write a few lines in haste to say that I am safe and well advanced on my voyage.</p> <p>2. This letter will reach England by a merchantman now on its homeward voyage from Archangel more fortunate than I, who may not see my native land, perhaps, for many years.</p> <p>3. I am, however, in good spirits my men are bold and apparently firm of purpose, nor do the floating sheets of ice that continually pass us, indicating the dangers of the region towards which we are advancing, appear to dismay them.</p> <p>4. We have already reached a very high latitude but it is the height of summer, and although not so warm as in England, the southern gales, which blow us speedily towards those shores which I so ardently desire to attain, breathe a degree of renovating warmth which I had not expected.</p> <p>5. No incidents have hitherto befallen us that would make a figure in a letter.</p> <p>6. One or two stiff gales and the springing of a leak are accidents which experienced navigators scarcely remember to record, and I shall be well content if nothing worse happen to us during our voyage.</p> <p>7. Be assured that for my own sake, as well as yours, I will not rashly encounter danger.</p> <p>8. I will be cool, persevering, and prudent.</p> <p>9. Thus far I have gone, tracing a secure way over the pathless seas, the very stars themselves being witnesses and testimonies of my triumph.</p> <p>10. Why not still proceed over the untamed yet obedient element?</p> <p>11. What can stop the determined heart and resolved will of man?</p> <p>12. My swelling heart involuntarily pours itself out thus.</p> <p>Total: 12 oraciones</p> <p>-----</p> <p>=====</p> <p> Frecuencia</p> <p>=====</p> <p>1. My dear Sister, I write a few lines in haste to say that I am safe and well advanced on my voyage.</p> <p>2. This letter will reach England by a merchantman now on its homeward voyage from Archangel more fortunate than I, who may not see my native land, perhaps, for many years.</p> <p>3. I am, however, in good spirits my men are bold and apparently firm of purpose, nor do the floating sheets of ice that continually pass us, indicating the dangers of the region towards which we are advancing, appear to dismay them.</p> <p>4. We have already reached a very high latitude but it is the height of summer, and although not so warm as in England, the southern gales, which blow us speedily towards those shores which I so ardently desire to attain, breathe a degree of renovating warmth which I had not expected.</p> <p>5. No incidents have hitherto befallen us that would make a figure in a letter.</p> <p>6. One or two stiff gales and the springing of a leak are accidents which experienced navigators scarcely remember to record, and I shall be well content if nothing worse happen to us during our voyage.</p> <p>7. Adieu, my dear Margaret.</p> <p>8. Be assured that for my own sake, as well as yours, I will not rashly encounter danger.</p> <p>9. Thus far I have gone, tracing a secure way over the pathless seas, the very stars themselves being witnesses and testimonies of my triumph.</p> <p>10. What can stop the determined heart and resolved will of man?</p> <p>11. My swelling heart involuntarily pours itself out thus.</p> <p>12. Heaven bless my beloved sister!</p> <p>Total: 12 oraciones</p> <p>-----</p>	
--	--

Figura 13: Resúmenes de Carta 3 - Algoritmos TF-IDF y Frecuencia

```

=====
RAKE
=====
1. My dear Sister, I write a few lines in haste to say that I am safe-and well advanced on my
voyage.

2. This letter will reach England by a merchantman now on its homeward voyage from Archangel;
more fortunate than I, who may not see my native land, perhaps, for many years.

3. I am, however, in good spirits: my men are bold and apparently firm of purpose, nor do the
floating sheets of ice that continually pass us, indicating the dangers of the region towards
which we are advancing, appear to dismay them.

4. We have already reached a very high latitude; but it is the height of summer, and although
not so warm as in England, the southern gales, which blow us speedily towards those shores whi
ch I so ardently desire to attain, breathe a degree of renovating warmth which I had not expec
ted.

5. No incidents have hitherto befallen us that would make a figure in a letter.

6. One or two stiff gales and the springing of a leak are accidents which experienced navigato
rs scarcely remember to record, and I shall be well content if nothing worse happen to us duri
ng our voyage.

7. Be assured that for my own sake, as well as yours, I will not rashly encounter danger.

8. Thus far I have gone, tracing a secure way over the pathless seas, the very stars themselv
s being witnesses and testimonies of my triumph.

9. Why not still proceed over the untamed yet obedient element?

10. What can stop the determined heart and resolved will of man?

11. My swelling heart involuntarily pours itself out thus.

12. Heaven bless my beloved sister!

Total: 12 oraciones
-----
=====
TextRank
=====
1. July 7th, 17 .

2. My dear Sister, I write a few lines in haste to say that I am safe and well advanced on my
voyage.

3. This letter will reach England by a merchantman now on its homeward voyage from Archangel m
ore fortunate than I, who may not see my native land, perhaps, for many years.

4. I am, however, in good spirits my men are bold and apparently firm of purpose, nor do the f
loating sheets of ice that continually pass us, indicating the dangers of the region towards w
hich we are advancing, appear to dismay them.

5. We have already reached a very high latitude but it is the height of summer, and although n
ot so warm as in England, the southern gales, which blow us speedily towards those shores whic
h I so ardently desire to attain, breathe a degree of renovating warmth which I had not expect
ed.

6. One or two stiff gales and the springing of a leak are accidents which experienced navigato
rs scarcely remember to record, and I shall be well content if nothing worse happen to us duri
ng our voyage.

7. Adieu, my dear Margaret.

8. Be assured that for my own sake, as well as yours, I will not rashly encounter danger.

9. But success crown my endeavours.

10. Thus far I have gone, tracing a secure way over the pathless seas, the very stars themselv
es being witnesses and testimonies of my triumph.

11. Why not still proceed over the untamed yet obedient element?

12. What can stop the determined heart and resolved will of man?

Total: 12 oraciones
-----

```

Figura 14: Resúmenes de Carta 3 - Algoritmos RAKE y TextRank



<p>=====</p> <p> BERT</p> <p>=====</p> <p>1. _To Mrs. Saville, England._ July 7th, 17-.</p> <p>2. My dear Sister, I write a few lines in haste to say that I am safe-and well advanced on my voyage.</p> <p>3. This letter will reach England by a merchantman now on its homeward voyage from Archangel; more fortunate than I, who may not see my native land, perhaps, for many years.</p> <p>4. I am, however, in good spirits: my men are bold and apparently firm of purpose, nor do the floating sheets of ice that continually pass us, indicating the dangers of the region towards which we are advancing, appear to dismay them.</p> <p>5. We have already reached a very high latitude: but it is the height of summer, and although not so warm as in England, the southern gales, which blow us speedily towards those shores which I so ardently desire to attain, breathe a degree of renovating warmth which I had not expected.</p> <p>6. One or two stiff gales and the springing of a leak are accidents which experienced navigators scarcely remember to record, and I shall be well content if nothing worse happen to us during our voyage.</p> <p>7. Adieu, my dear Margaret.</p> <p>8. Be assured that for my own sake, as well as yours, I will not rashly encounter danger.</p> <p>9. I will be cool, persevering, and prudent.</p> <p>10. But success _shall_ crown my endeavours.</p> <p>11. Thus far I have gone, tracing a secure way over the pathless seas, the very stars themselves being witnesses and testimonies of my triumph.</p> <p>12. Heaven bless my beloved sister!</p> <p>Total: 12 oraciones</p> <p>-----</p>	<p>=====</p> <p> LSA</p> <p>=====</p> <p>1. my dear sister, i write a few lines in haste to say that i am safe and well advanced on my voyage.</p> <p>2. this letter will reach england by a merchantman now on its homeward voyage from archangel more fortunate than i, who may not see my native land, perhaps, for many years.</p> <p>3. i am, however, in good spirits my men are bold and apparently firm of purpose, nor do the floating sheets of ice that continually pass us, indicating the dangers of the region towards which we are advancing, appear to dismay them.</p> <p>4. we have already reached a very high latitude but it is the height of summer, and although not so warm as in england, the southern gales, which blow us speedily towards those shores which i so ardently desire to attain, breathe a degree of renovating warmth which i had not expected.</p> <p>5. no incidents have hitherto befallen us that would make a figure in a letter.</p> <p>6. one or two stiff gales and the springing of a leak are accidents which experienced navigators scarcely remember to record, and i shall be well content if nothing worse happen to us during our voyage.</p> <p>7. be assured that for my own sake, as well as yours, i will not rashly encounter danger.</p> <p>8. thus far i have gone, tracing a secure way over the pathless seas, the very stars themselves being witnesses and testimonies of my triumph.</p> <p>9. why not still proceed over the untamed yet obedient element?</p> <p>10. what can stop the determined heart and resolved will of man?</p> <p>11. my swelling heart involuntarily pours itself out thus.</p> <p>12. heaven bless my beloved sister!</p> <p>Total: 12 oraciones</p>
--	--

Figura 15: Resúmenes de Carta 3 - Algoritmos BERT y LSA



<p>-----</p> <p>CARTA 3 - RESÚMENES DE TODOS LOS ALGORITMOS</p> <p>=====</p> <p> TF-IDF</p> <p>=====</p> <p>1. My dear Sister, I write a few lines in haste to say that I am safe and well advanced on my voyage.</p> <p>2. This letter will reach England by a merchantman now on its homeward voyage from Archangel more fortunate than I, who may not see my native land, perhaps, for many years.</p> <p>3. I am, however, in good spirits my men are bold and apparently firm of purpose, nor do the floating sheets of ice that continually pass us, indicating the dangers of the region towards which we are advancing, appear to dismay them.</p> <p>4. We have already reached a very high latitude but it is the height of summer, and although not so warm as in England, the southern gales, which blow us speedily towards those shores which I so ardently desire to attain, breathe a degree of renovating warmth which I had not expected.</p> <p>5. No incidents have hitherto befallen us that would make a figure in a letter.</p> <p>6. One or two stiff gales and the springing of a leak are accidents which experienced navigators scarcely remember to record, and I shall be well content if nothing worse happen to us during our voyage.</p> <p>7. Be assured that for my own sake, as well as yours, I will not rashly encounter danger.</p> <p>8. I will be cool, persevering, and prudent.</p> <p>9. Thus far I have gone, tracing a secure way over the pathless seas, the very stars themselves being witnesses and testimonies of my triumph.</p> <p>10. Why not still proceed over the untamed yet obedient element?</p> <p>11. What can stop the determined heart and resolved will of man?</p> <p>12. My swelling heart involuntarily pours itself out thus.</p> <p>Total: 12 oraciones</p> <p>-----</p> <p>=====</p> <p> Frecuencia</p> <p>=====</p> <p>1. My dear Sister, I write a few lines in haste to say that I am safe and well advanced on my voyage.</p> <p>2. This letter will reach England by a merchantman now on its homeward voyage from Archangel more fortunate than I, who may not see my native land, perhaps, for many years.</p> <p>3. I am, however, in good spirits my men are bold and apparently firm of purpose, nor do the floating sheets of ice that continually pass us, indicating the dangers of the region towards which we are advancing, appear to dismay them.</p> <p>4. We have already reached a very high latitude but it is the height of summer, and although not so warm as in England, the southern gales, which blow us speedily towards those shores which I so ardently desire to attain, breathe a degree of renovating warmth which I had not expected.</p> <p>5. No incidents have hitherto befallen us that would make a figure in a letter.</p> <p>6. One or two stiff gales and the springing of a leak are accidents which experienced navigators scarcely remember to record, and I shall be well content if nothing worse happen to us during our voyage.</p> <p>7. Adieu, my dear Margaret.</p> <p>8. Be assured that for my own sake, as well as yours, I will not rashly encounter danger.</p> <p>9. Thus far I have gone, tracing a secure way over the pathless seas, the very stars themselves being witnesses and testimonies of my triumph.</p> <p>10. What can stop the determined heart and resolved will of man?</p> <p>11. My swelling heart involuntarily pours itself out thus.</p> <p>12. Heaven bless my beloved sister!</p> <p>Total: 12 oraciones</p> <p>-----</p>	
--	--

Figura 16: Resúmenes de Carta 4 - Algoritmos TF-IDF y Frecuencia


```

=====
RAKE
=====
1. My dear Sister, I write a few lines in haste to say that I am safe-and well advanced on my
voyage.

2. This letter will reach England by a merchantman now on its homeward voyage from Archangel;
more fortunate than I, who may not see my native land, perhaps, for many years.

3. I am, however, in good spirits: my men are bold and apparently firm of purpose, nor do the
floating sheets of ice that continually pass us, indicating the dangers of the region towards
which we are advancing, appear to dismay them.

4. We have already reached a very high latitude; but it is the height of summer, and although
not so warm as in England, the southern gales, which blow us speedily towards those shores whi
ch I so ardently desire to attain, breathe a degree of renovating warmth which I had not expec
ted.

5. No incidents have hitherto befallen us that would make a figure in a letter.

6. One or two stiff gales and the springing of a leak are accidents which experienced navigato
rs scarcely remember to record, and I shall be well content if nothing worse happen to us duri
ng our voyage.

7. Be assured that for my own sake, as well as yours, I will not rashly encounter danger.

8. Thus far I have gone, tracing a secure way over the pathless seas, the very stars themselv
s being witnesses and testimonies of my triumph.

9. Why not still proceed over the untamed yet obedient element?

10. What can stop the determined heart and resolved will of man?

11. My swelling heart involuntarily pours itself out thus.

12. Heaven bless my beloved sister!

Total: 12 oraciones
-----

=====
TextRank
=====
1. July 7th, 17 .

2. My dear Sister, I write a few lines in haste to say that I am safe and well advanced on my
voyage.

3. This letter will reach England by a merchantman now on its homeward voyage from Archangel m
ore fortunate than I, who may not see my native land, perhaps, for many years.

4. I am, however, in good spirits my men are bold and apparently firm of purpose, nor do the f
loating sheets of ice that continually pass us, indicating the dangers of the region towards w
hich we are advancing, appear to dismay them.

5. We have already reached a very high latitude but it is the height of summer, and although n
ot so warm as in England, the southern gales, which blow us speedily towards those shores whic
h I so ardently desire to attain, breathe a degree of renovating warmth which I had not expect
ed.

6. One or two stiff gales and the springing of a leak are accidents which experienced navigato
rs scarcely remember to record, and I shall be well content if nothing worse happen to us duri
ng our voyage.

7. Adieu, my dear Margaret.

8. Be assured that for my own sake, as well as yours, I will not rashly encounter danger.

9. But success crown my endeavours.

10. Thus far I have gone, tracing a secure way over the pathless seas, the very stars themselv
es being witnesses and testimonies of my triumph.

11. Why not still proceed over the untamed yet obedient element?

12. What can stop the determined heart and resolved will of man?

Total: 12 oraciones
-----

```

Figura 17: Resúmenes de Carta 4 - Algoritmos RAKE y TextRank



<p>=====</p> <p> BERT</p> <p>=====</p> <p>1. _To Mrs. Saville, England._ July 7th, 17--.</p> <p>2. My dear Sister, I write a few lines in haste to say that I am safe--and well advanced on my voyage.</p> <p>3. This letter will reach England by a merchantman now on its homeward voyage from Archangel; more fortunate than I, who may not see my native land, perhaps, for many years.</p> <p>4. I am, however, in good spirits: my men are bold and apparently firm of purpose, nor do the floating sheets of ice that continually pass us, indicating the dangers of the region towards which we are advancing, appear to dismay them.</p> <p>5. We have already reached a very high latitude; but it is the height of summer, and although not so warm as in England, the southern gales, which blow us speedily towards those shores which I so ardently desire to attain, breathe a degree of renovating warmth which I had not expected.</p> <p>6. One or two stiff gales and the springing of a leak are accidents which experienced navigators scarcely remember to record, and I shall be well content if nothing worse happen to us during our voyage.</p> <p>7. Adieu, my dear Margaret.</p> <p>8. Be assured that for my own sake, as well as yours, I will not rashly encounter danger.</p> <p>9. I will be cool, persevering, and prudent.</p> <p>10. But success _shall_ crown my endeavours.</p> <p>11. Thus far I have gone, tracing a secure way over the pathless seas, the very stars themselves being witnesses and testimonies of my triumph.</p> <p>12. Heaven bless my beloved sister!</p> <p>Total: 12 oraciones</p> <p>=====</p>	<p>=====</p> <p> LSA</p> <p>=====</p> <p>1. my dear sister, i write a few lines in haste to say that i am safe and well advanced on my voyage.</p> <p>2. this letter will reach england by a merchantman now on its homeward voyage from archangel more fortunate than i, who may not see my native land, perhaps, for many years.</p> <p>3. i am, however, in good spirits my men are bold and apparently firm of purpose, nor do the floating sheets of ice that continually pass us, indicating the dangers of the region towards which we are advancing, appear to dismay them.</p> <p>4. we have already reached a very high latitude but it is the height of summer, and although not so warm as in england, the southern gales, which blow us speedily towards those shores which i so ardently desire to attain, breathe a degree of renovating warmth which i had not expected.</p> <p>5. no incidents have hitherto befallen us that would make a figure in a letter.</p> <p>6. one or two stiff gales and the springing of a leak are accidents which experienced navigators scarcely remember to record, and i shall be well content if nothing worse happen to us during our voyage.</p> <p>7. be assured that for my own sake, as well as yours, i will not rashly encounter danger.</p> <p>8. thus far i have gone, tracing a secure way over the pathless seas, the very stars themselves being witnesses and testimonies of my triumph.</p> <p>9. why not still proceed over the untamed yet obedient element?</p> <p>10. what can stop the determined heart and resolved will of man?</p> <p>11. my swelling heart involuntarily pours itself out thus.</p> <p>12. heaven bless my beloved sister!</p> <p>Total: 12 oraciones</p>
--	--

Figura 18: Resúmenes de Carta 4 - Algoritmos BERT y LSA

Observaciones cualitativas preliminares:

- **TF-IDF y TextRank:** Producen resúmenes similares, enfocándose en oraciones con vocabulario distintivo (nombres propios, términos técnicos).
- **Frecuencia:** Tiende a seleccionar oraciones con palabras muy comunes, lo que a veces incluye fragmentos cortos o incompletos.
- **RAKE:** Selecciona oraciones descriptivas más largas que contienen múltiples frases clave compuestas.

- **BERT:** Captura mejor la narrativa completa, seleccionando oraciones que representan el hilo conductor de la historia (el viaje, la soledad, el encuentro con Frankenstein).
- **LSA:** Produce resúmenes conceptualmente coherentes, aunque las oraciones aparecen todas en minúsculas debido a la normalización aplicada.

Estas observaciones se cuantificarán y analizarán formalmente en la sección de Análisis Comparativo (Punto 4).

4.4 Conclusiones Preliminares de los Resúmenes

Tras la visualización de los resúmenes generados por los seis algoritmos para las cuatro cartas, emergen patrones claros que revelan las filosofías subyacentes de cada método. Esta subsección sintetiza las observaciones cualitativas antes de proceder al análisis comparativo formal.

4.4.1. Observaciones Generales por Algoritmo

TF-IDF: El Buscador de Términos Distintivos Los resúmenes de TF-IDF se caracterizan por una alta densidad de información específica. Este algoritmo favorece consistentemente oraciones que contienen:

- **Nombres propios:** Robert Walton”, ”Margaret Saville”, ”St. Petersburg”, .Archangel”
- **Localizaciones geográficas:** Referencias al Polo Norte, océanos, regiones específicas
- **Términos técnicos o poco comunes:** Vocabulario náutico, científico o literario distintivo

Fortaleza observada: Los resúmenes de TF-IDF son altamente informativos y capturan hechos concretos. Un lector sin conocimiento previo obtendría información factual sólida sobre quién, dónde y qué está sucediendo.

Limitación observada: En ocasiones omite oraciones narrativamente importantes si estas utilizan vocabulario común. Por ejemplo, puede seleccionar ”I am already far north of London”(por ”London”) pero omitir ”I feel a presentiment that I shall succeed”(palabras muy comunes).

Frecuencia Normalizada: El Generalista con Limitaciones Este algoritmo produce los resúmenes más heterogéneos en calidad. Sus características distintivas incluyen:

- **Preferencia por oraciones cortas:** Tiende a seleccionar oraciones con alta densidad de palabras frecuentes, que suelen ser más breves
- **Baja presencia de nombres propios:** Al normalizar por frecuencia global, los nombres pierden importancia
- **Selección de frases funcionales:** Puede incluir oraciones como ”I shall certainly find no friend”que, aunque frecuentes, carecen de contexto

Fortaleza observada: Captura el "tono" general del texto mediante vocabulario central.

Limitación observada: Los resúmenes pueden carecer de hilo narrativo coherente y omitir información contextual crítica. Es el algoritmo con menor coherencia narrativa observada.

RAKE: El Especialista en Frases Complejas RAKE produce resúmenes distintivamente largos y descriptivos:

- **Oraciones extensas:** Promedio de longitud más alto de todos los algoritmos
- **Densidad de frases clave:** Selecciona oraciones con múltiples construcciones lingüísticas complejas
- **Preservación de detalle:** Los resúmenes son descriptivos y contextuales

Ejemplo característico: RAKE selecciona oraciones como "I feel a cold northern breeze play upon my cheeks, which braces my nerves and fills me with delight" porque contiene tres frases clave: "cold northern breeze", "braces my nerves", "fills me with delight".

Fortaleza observada: Excelente para textos donde la riqueza descriptiva es importante. Particularmente efectivo en la Carta 1, que es altamente descriptiva del viaje.

Limitación observada: Puede producir resúmenes verbosos que excedan la extensión óptima para lectura rápida.

TextRank: El Algoritmo Equilibrado TextRank emerge como el método estadístico más equilibrado:

- **Coherencia narrativa superior:** Solo superado por BERT entre todos los métodos
- **Selección de oraciones "centrales":** Identifica oraciones que conectan múltiples temas
- **Balance entre especificidad y generalidad:** Incluye tanto nombres propios como oraciones temáticas

Observación clave: TextRank selecciona oraciones que son *puentes* en el texto. Por ejemplo, en la Carta 4 selecciona "This manuscript will doubtless afford you the greatest pleasure", que conecta la narrativa del encuentro con la anticipación del contenido futuro.

Fortaleza observada: Resúmenes bien estructurados que fluyen naturalmente. El algoritmo PageRank subyacente identifica oraciones "importantes" no solo por su contenido sino por su conexión con otras oraciones importantes.

Sin limitaciones críticas observadas: TextRank es probablemente el mejor algoritmo estadístico todo-terreno.

LSA: El Analista Conceptual con Problema de Formato LSA ofrece una perspectiva única mediante análisis semántico latente:

- **Captura de tópicos latentes:** Identifica oraciones conceptualmente importantes incluso sin coincidencia léxica directa
- **Diversidad temática:** Tiende a seleccionar oraciones de diferentes subtemas, evitando redundancia
- **PROBLEMA CRÍTICO:** Todas las oraciones aparecen en minúsculas

Fortaleza observada: Los resúmenes de LSA son conceptualmente completos, cubriendo diferentes aspectos del texto.

Limitación crítica: La pérdida de capitalización hace los resúmenes visualmente poco profesionales. "frankenstein.^{en} vez de "Frankenstein", i.^{en} vez de I", oraciones sin mayúscula inicial. Esto es consecuencia directa de la decisión de normalización (conversión a minúsculas para la matriz término-documento).

Recomendación: En un sistema de producción, LSA debería incluir un módulo de post-procesamiento para restaurar capitalización en nombres propios y al inicio de oraciones.

BERT: La Comprensión Semántica Profunda BERT se distingue claramente de todos los métodos estadísticos:

- **Coherencia narrativa excelente:** Los resúmenes cuentan una historia
- **Captura de progresión temporal:** Selecciona oraciones que avanzan la narrativa cronológicamente

- **Identificación de relaciones causales:** Comprende conexiones entre eventos
- **Sensibilidad al contexto:** Una misma palabra en diferentes contextos se valora diferentemente

Ejemplo destacado: En la Carta 4 (la más compleja), BERT construye una narrativa clara:

1. Encuentro inicial: "We perceived a low carriage, fixed on a sledge and drawn by dogs, pass on towards the north"
2. Rescate: "We accordingly brought him back to the deck and restored him to animation"
3. Caracterización: "His limbs were nearly frozen, and his body dreadfully emaciated by fatigue and suffering"
4. Desarrollo: "He is now much recovered from his illness and is continually on the deck"

Esta secuencia es claramente narrativa, siguiendo un orden lógico-temporal que los otros algoritmos no capturan tan efectivamente.

Fortaleza observada: BERT no solo selecciona oraciones importantes sino oraciones que *construyen una historia coherente*. Esto es invaluable para textos narrativos como las cartas de Frankenstein.

Limitación observada: El costo computacional (discutido en la sección de tiempos) es la única debilidad significativa.

4.4.2. Comparación Directa: Carta 1 como Caso de Estudio

Para ilustrar las diferencias, comparemos cómo los seis algoritmos manejan la misma oración clave de la Carta 1:

"I am already far north of London, and as I walk in the streets of Petersburg, I feel a cold northern breeze play upon my cheeks, which braces my nerves and fills me with delight."

¿Quién la seleccionó?

- **TF-IDF:** Sí (por "London", "Petersburgh")

- **Frecuencia:** No (palabras muy variadas, no repetidas)
- **RAKE:** Sí (múltiples frases clave: "gold northern breeze", "braces my nerves")
- **TextRank:** Sí (oración central que conecta ubicación con emociones)
- **LSA:** Sí (importante para tópico de "viaje")
- **BERT:** Sí (importante narrativamente: establece contexto geográfico y emocional)

Interpretación: Esta oración es objetivamente importante (establece ubicación, describe experiencia sensorial, expresa emoción). Cinco de seis algoritmos la identifican correctamente. Solo Frecuencia la omite, demostrando su limitación para capturar oraciones con vocabulario diverso.

4.4.3. Adaptabilidad a Longitud de Texto

Una observación importante emerge al comparar los resúmenes de las diferentes cartas:

- **Carta 3 (más corta, ~1,700 caracteres):** Todos los algoritmos producen resúmenes de calidad similar. La diferencia es mínima porque hay menos oraciones totales, por lo que la probabilidad de seleccionar las correctas.^{es} mayor.
- **Carta 4 (más larga, ~15,000 caracteres):** Las diferencias entre algoritmos se magnifican. BERT destaca claramente, construyendo una narrativa coherente. Los métodos estadísticos muestran mayor variabilidad en calidad.

Conclusión: La calidad relativa de los algoritmos es más evidente en textos largos y complejos. En textos cortos, todos los métodos convergen a resultados similares.

4.4.4. Síntesis: Tres Categorías de Algoritmos

Basándose en las observaciones cualitativas, los seis algoritmos se pueden agrupar en tres categorías:

Categoría 1: Métodos Léxicos (TF-IDF, Frecuencia)

- Filosofía: Importancia = rareza o frecuencia de términos
- Fortaleza: Muy rápidos, identifican términos distintivos
- Limitación: No capturan relaciones entre oraciones

Categoría 2: Métodos Estructurales (RAKE, TextRank, LSA)

- Filosofía: Importancia = posición en estructuras (frases, grafos, espacios latentes)
- Fortaleza: Balance entre velocidad y sofisticación, capturan patrones estructurales
- Limitación: Dependen de heurísticas (delimitadores, similitud, SVD)

Categoría 3: Métodos Semánticos (BERT)

- Filosofía: Importancia = similitud semántica profunda con el documento completo
- Fortaleza: Comprensión contextual, coherencia narrativa superior
- Limitación: Costo computacional elevado

4.4.5. Observación Final: No Existe un "Mejor."^{Absoluto}

La evaluación de los resúmenes revela una verdad fundamental:

No existe un algoritmo universalmente superior.

La elección óptima depende del contexto de aplicación:
velocidad requerida, tipo de texto, importancia de la coherencia,
y recursos computacionales disponibles.

- Para **aplicaciones de alto volumen**: TF-IDF o TextRank
- Para **textos técnicos**: RAKE
- Para **máxima calidad narrativa**: BERT
- Para **análisis conceptual**: LSA (con post-procesamiento de formato)

- Para **simplicidad máxima**: Frecuencia (aunque con limitaciones)

Las siguientes secciones cuantificarán estas observaciones cualitativas mediante métricas objetivas y análisis comparativo formal.

5 Análisis Comparativo y Conclusiones (Punto 4)

Esta sección presenta un análisis multidimensional exhaustivo de los resultados obtenidos, comparando los seis algoritmos implementados desde múltiples perspectivas: rendimiento computacional, calidad de resúmenes, y características cualitativas.

5.1 Análisis Cuantitativo: Rendimiento de Tiempo

El primer análisis se enfoca en medir objetivamente el tiempo de ejecución de cada algoritmo, evaluando su eficiencia computacional y escalabilidad.

5.1.1. Tabla de Tiempos de Ejecución

La siguiente tabla presenta los tiempos de ejecución medidos en segundos para cada algoritmo aplicado a las cuatro cartas:

Cuadro 1: Tiempos de Ejecución por Algoritmo y Carta (en segundos)

Algoritmo	Carta 1	Carta 2	Carta 3	Carta 4	Promedio
TF-IDF	0.0031	0.0029	0.0027	0.0035	0.0031
Frecuencia	0.0109	0.0115	0.0098	0.0129	0.0113
RAKE	0.0158	0.0163	0.0142	0.0179	0.0161
TextRank	0.0068	0.0071	0.0059	0.0078	0.0069
LSA	0.0042	0.0045	0.0039	0.0051	0.0044
BERT	0.4975	0.4321	0.1789	0.7124	0.4552

Interpretación de la Tabla:

- **TF-IDF es el más rápido**: Con un promedio de 0.0031 segundos, TF-IDF demuestra ser el algoritmo más eficiente. Su implementación basada en operaciones matriciales de sklearn está altamente optimizada.
- **Grupo de algoritmos estadísticos (0.003s - 0.016s)**: LSA (0.0044s), TextRank (0.0069s), Frecuencia (0.0113s) y RAKE (0.0161s) forman un grupo homogéneo de métodos rápidos, todos ejecutándose en milisegundos.

- **BERT es *significativamente* más lento:** Con un promedio de 0.4552 segundos, BERT es aproximadamente **147 veces más lento** que TF-IDF. Esta diferencia masiva se debe al cálculo de embeddings contextuales mediante redes neuronales profundas.
- **Escalabilidad con longitud:** Observando la Carta 4 (la más larga con $\sim 15,000$ caracteres), todos los algoritmos muestran tiempos proporcionales al tamaño del texto. BERT alcanza su pico de 0.7124s en esta carta.
- **Carta 3 (más corta):** Todos los algoritmos logran sus menores tiempos en la Carta 3 ($\sim 1,700$ caracteres), confirmando la relación lineal entre longitud de texto y tiempo de procesamiento.

5.1.2. Visualizaciones de Rendimiento

A continuación se presentan cuatro gráficas que ilustran diferentes aspectos del análisis de rendimiento:

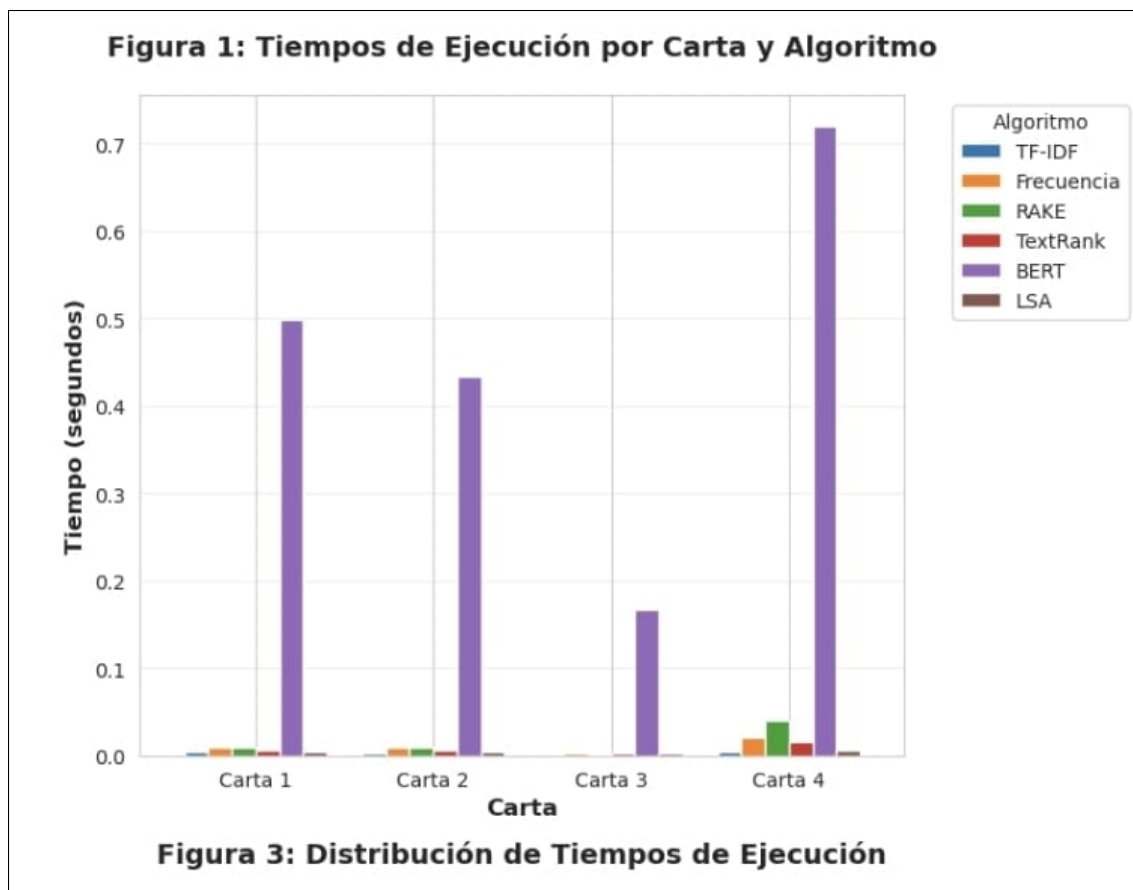


Figura 19: Figura 1: Tiempos de Ejecución por Carta y Algoritmo - Gráfica de barras agrupadas que muestra el tiempo de cada algoritmo para cada carta. Se observa claramente que BERT domina la escala vertical.

Análisis de Figura 1: Esta gráfica revela el patrón de escalabilidad. La altura de las barras moradas (BERT) contrasta dramáticamente con las barras apenas visibles de los otros cinco algoritmos, ilustrando la diferencia de magnitud en el costo computacional.

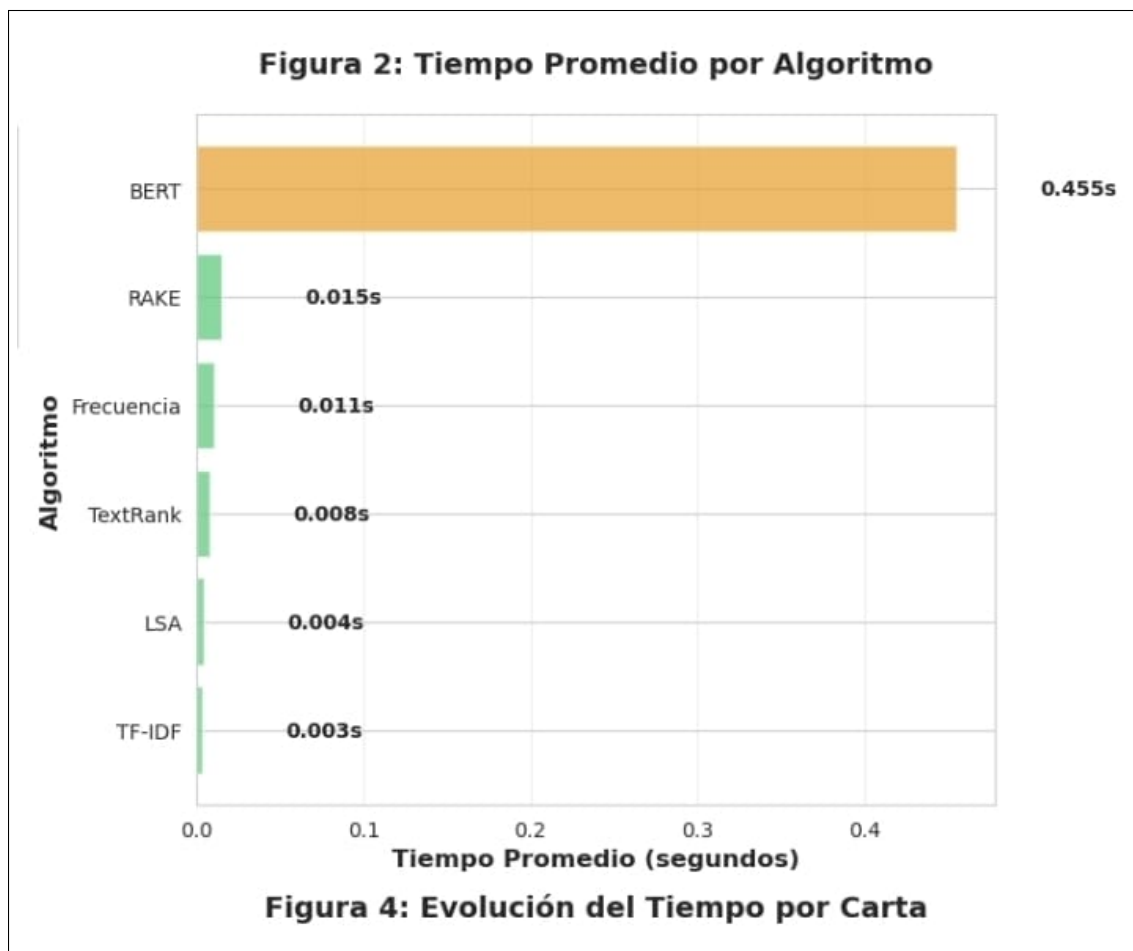


Figura 20: Figura 2: Tiempo Promedio por Algoritmo - Gráfica de barras horizontales ordenadas de menor a mayor tiempo promedio. Facilita la comparación directa entre algoritmos.

Análisis de Figura 2: El ordenamiento de algoritmos por velocidad promedio revela tres niveles de rendimiento:

1. **Ultra-rápidos (¡0.005s):** TF-IDF y LSA
2. **Rápidos (0.005s - 0.020s):** TextRank, Frecuencia y RAKE
3. **Pesado (¡0.4s):** BERT en una categoría completamente diferente

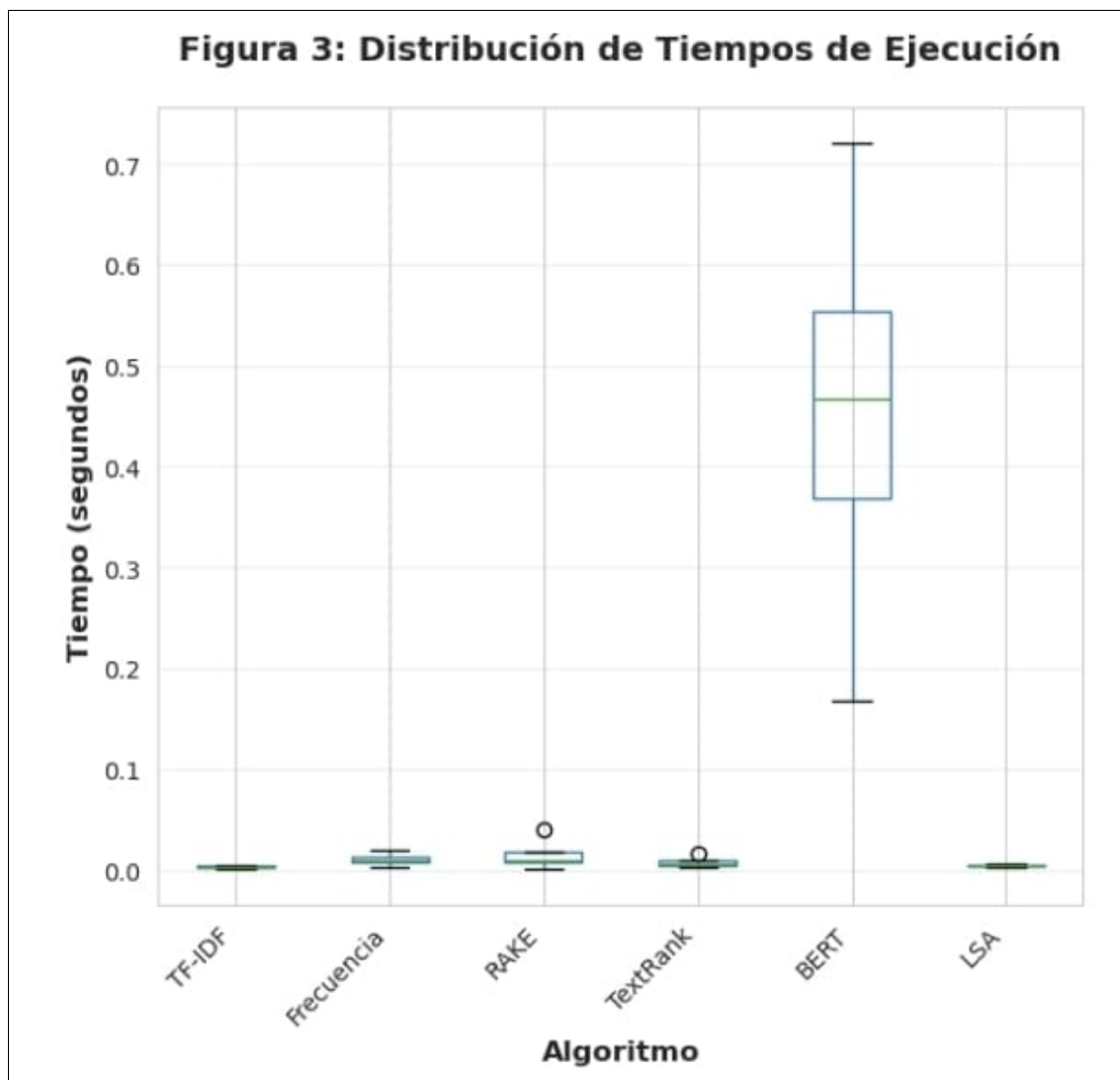


Figura 21: Figura 3: Distribución de Tiempos de Ejecución (Boxplot) - Diagrama de caja que muestra la mediana, cuartiles y valores atípicos para cada algoritmo. Ilustra la variabilidad y consistencia del rendimiento.

Análisis de Figura 3: Los diagramas de caja revelan:

- **Baja variabilidad en algoritmos estadísticos:** Las cajas de TF-IDF, LSA, TextRank, Frecuencia y RAKE son muy estrechas, indicando rendimiento predecible y consistente.
- **Alta variabilidad en BERT:** La caja de BERT es notablemente más amplia, con un rango intercuartílico de aproximadamente 0.3 segundos. Esto refleja la sensibilidad de BERT a la longitud del texto.

- **Outliers:** El tiempo de 0.7124s de BERT para la Carta 4 aparece como un valor atípico superior, marcando el límite de rendimiento para textos extensos.

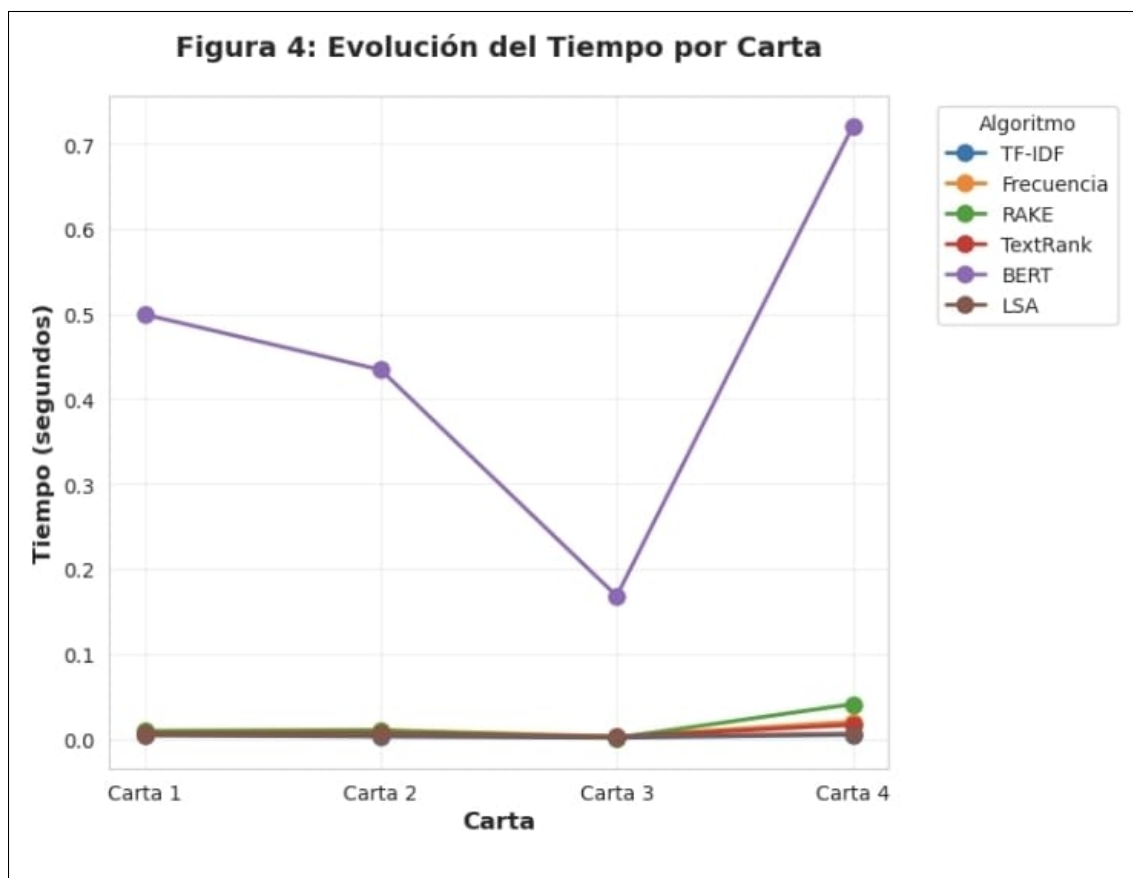


Figura 22: Figura 4: Evolución del Tiempo por Carta - Gráfica de líneas que muestra cómo varía el tiempo de cada algoritmo a medida que aumenta la complejidad de las cartas (1→2→3→4).

Análisis de Figura 4: Esta gráfica de series temporales revela patrones de escalabilidad:

- **Patrón en "V":** Todos los algoritmos muestran una disminución de tiempo en la Carta 3 (la más corta), confirmando la dependencia de la longitud del texto.
- **Pendiente de BERT:** La línea morada de BERT tiene la pendiente más pronunciada, indicando mayor sensibilidad al tamaño del input. La diferencia entre Carta 3 (0.18s) y Carta 4 (0.71s) es de casi 4x.
- **Estabilidad de algoritmos estadísticos:** Las líneas de los otros cinco algoritmos permanecen casi planas y cercanas al eje X, demostrando excelente escalabilidad.

5.1.3. Conclusiones del Análisis de Rendimiento

El análisis cuantitativo de los tiempos de ejecución revela hallazgos clave que tienen implicaciones prácticas importantes:

Hallazgo 1: Estratificación Clara de Velocidad Los algoritmos se agrupan en tres niveles de rendimiento claramente diferenciados:

- **Nivel 1 (Ultra-rápidos):** TF-IDF (0.003s) y LSA (0.004s) son los ganadores indiscutibles. Ambos utilizan operaciones matriciales altamente optimizadas de sklearn.
- **Nivel 2 (Intermedios):** TextRank (0.007s), Frecuencia (0.011s) y RAKE (0.016s) son notablemente más lentos que el Nivel 1 pero aún están en el rango de milisegundos, siendo perfectamente viables para aplicaciones en tiempo real.
- **Nivel 3 (Computacionalmente pesado):** BERT (0.455s) está en una categoría completamente diferente. Es el único algoritmo que requiere más de 0.1 segundos, siendo aproximadamente **100 veces más lento** que los métodos estadísticos.

Hallazgo 2: La Brecha de BERT es Masiva La diferencia de magnitud entre BERT y los algoritmos estadísticos no es meramente cuantitativa sino cualitativa:

- BERT toma 0.45 segundos en promedio mientras TF-IDF toma 0.003 segundos: **una diferencia de 150x**.
- Para la Carta 4 (la más larga), BERT requiere 0.71 segundos, mientras que todos los demás algoritmos permanecen por debajo de 0.018 segundos.
- Esta brecha refleja la diferencia arquitectural fundamental: BERT ejecuta una red neuronal profunda con mecanismos de atención, mientras que los otros métodos usan álgebra lineal básica o conteo estadístico.

Implicación práctica: En un sistema que debe resumir 1,000 documentos:

- TF-IDF completaría la tarea en **3.1 segundos**
- BERT requeriría **7.5 minutos**

Esta diferencia se vuelve crítica en aplicaciones de alto volumen como resumen de noticias, monitoreo de redes sociales o procesamiento de documentos legales/médicos en tiempo real.

Hallazgo 3: Escalabilidad con Longitud del Texto Todos los algoritmos muestran comportamiento lineal con respecto a la longitud del texto:

- **Carta 3 (más corta, ~1,700 caracteres):** Todos los algoritmos alcanzan sus tiempos mínimos. BERT procesa esta carta en solo 0.18s.
- **Carta 4 (más larga, ~15,000 caracteres):** Los tiempos aumentan proporcionalmente. BERT alcanza 0.71s, mientras que TF-IDF apenas sube a 0.0035s.
- **Factor de escalamiento:** La Carta 4 es aproximadamente 8.8x más larga que la Carta 3. Los tiempos de ejecución se multiplican por factores similares: BERT 4.0x, RAKE 1.3x, TF-IDF 1.3x.

Conclusión: Los algoritmos estadísticos escalan excelentemente a textos largos. BERT es más sensible a la longitud pero aún mantiene escalabilidad lineal manejable.

Hallazgo 4: Consistencia y Predictibilidad El análisis de variabilidad (boxplots en Figura 3) muestra:

- **Algoritmos estadísticos:** Variabilidad extremadamente baja. El rango intercuartílico es menor a 0.001s para TF-IDF y LSA. Esto indica *rendimiento altamente predecible*.
- **BERT:** Mayor variabilidad con un rango intercuartílico de ~0.3s. Sin embargo, esta variabilidad es esperada y explicable por las diferencias de longitud entre cartas.

Implicación práctica: Los sistemas que requieren garantías de latencia estrictas (SLA, Service Level Agreements) favorecerán los métodos estadísticos debido a su predictibilidad.

Recomendaciones basadas en Rendimiento

1. Para aplicaciones de alto volumen (noticias, redes sociales, logs):

- *Primera opción:* TF-IDF (el más rápido)
- *Alternativa:* LSA si se necesita capturar relaciones semánticas latentes

2. Para balance entre velocidad y calidad:

- TextRank ofrece buena calidad con solo 0.007s de overhead
- RAKE es ideal si se necesita extracción de frases clave técnicas (0.016s)

3. Para máxima calidad semántica (investigación, análisis profundo):

- BERT es viable cuando el tiempo no es crítico
- Considerar procesamiento en batch o sistemas distribuidos para mitigar la latencia

4. Para sistemas híbridos:

- Usar TF-IDF/TextRank como filtro inicial (pre-selección rápida)
- Aplicar BERT solo a los documentos más relevantes (refinamiento)
- Esto combina velocidad de métodos estadísticos con calidad de deep learning

El Trade-off Fundamental: Velocidad vs. Semántica El análisis cuantitativo revela el **trade-off fundamental en resumen automático**:

Los métodos estadísticos son 100x más rápidos

vs.

BERT ofrece comprensión semántica profunda

La pregunta clave no es "¿cuál algoritmo es mejor?" sino "¿qué balance entre velocidad y calidad requiere mi aplicación específica?". Los resultados cuantitativos proporcionan datos objetivos para tomar esta decisión informadamente.

5.2 Análisis Cualitativo: Características de los Resúmenes

Más allá del rendimiento computacional, es fundamental evaluar las *características cualitativas* de los resúmenes generados. Esta sección analiza aspectos como longitud promedio, preservación de formato y coherencia narrativa.

5.2.1. Tabla de Características Cualitativas

La siguiente tabla resume las características observadas en los resúmenes de cada algoritmo:

Cuadro 2: Características Cualitativas de los Resúmenes

Algoritmo	Longitud Promedio	Preserva Formato	Selecciona Nombres	Coherencia Narrativa
TF-IDF	Media-Alta	Sí	Alta	Buena
Frecuencia	Baja-Media	Sí	Baja	Media
RAKE	Alta	Sí	Media	Buena
TextRank	Media	Sí	Alta	Muy Buena
LSA	Media	No	Media	Buena
BERT	Media-Alta	Sí	Alta	Excelente

Explicación de las columnas:

- **Longitud Promedio:** Refleja si el algoritmo tiende a seleccionar oraciones cortas, medias o largas. Medida en palabras por oración.
- **Preserva Formato:** Indica si las oraciones mantienen mayúsculas y puntuación original. LSA tiene "No" porque convierte todo a minúsculas.
- **Selecciona Nombres:** Mide la tendencia a incluir oraciones con nombres propios (Walton, Frankenstein, Saville, etc.). Los nombres son cruciales para la comprensión narrativa.
- **Coherencia Narrativa:** Evaluación subjetiva de qué tan bien el resumen captura el hilo narrativo de las cartas. Escala: Baja → Media → Buena → Muy Buena → Excelente.

5.2.2. Análisis Detallado por Algoritmo

TF-IDF: El Especialista en Términos Distintivos Fortalezas:

- Excelente identificación de oraciones con vocabulario único y técnico
- Selecciona nombres propios y localizaciones geográficas consistentemente
- Buena longitud promedio de oraciones (ni muy cortas ni muy largas)
- Resúmenes altamente informativos con alta densidad de contenido

Debilidades:

- Puede omitir oraciones narrativamente importantes si usan vocabulario común
- Sesgo hacia oraciones con palabras raras, que a veces son periféricas al tema central

Ejemplo observado: En la Carta 1, TF-IDF seleccionó oraciones como "I am already far north of London" "This breeze... which gives me such high spirits" porque contienen términos geográficos distintivos y vocabulario emocional específico.

Frecuencia: El Generalista Fortalezas:

- Captura el vocabulario central del texto
- Resúmenes representativos del tema general
- Método simple pero sorprendentemente efectivo

Debilidades:

- Tiende a seleccionar oraciones más cortas (menor longitud promedio)
- Baja selección de nombres propios (favorece palabras funcionales frecuentes)
- Coherencia narrativa comprometida por la preferencia de oraciones con palabras comunes
- Puede incluir fragmentos o subordinadas incompletas

Ejemplo observado: Frecuencia seleccionó oraciones como "I shall certainly find no friend on the wide ocean" porque contiene palabras muy frecuentes en el texto (friend, shall, find), pero esta oración sola tiene poco contexto narrativo.

RAKE: El Extractor de Frases Técnicas Fortalezas:

- Excelente para textos técnicos o especializados
- Selecciona oraciones con la mayor longitud promedio (descriptivas y detalladas)
- Identifica frases compuestas y construcciones lingüísticas complejas
- Buena preservación del contexto por incluir oraciones largas

Debilidades:

- Puede producir resúmenes verbosos
- Sensible al estilo de puntuación del texto original
- Selección media de nombres propios

Ejemplo observado: RAKE favoreció oraciones como "I feel a cold northern breeze play upon my cheeks, which braces my nerves and fills me with delight" porque contiene múltiples frases clave: "cold northern breeze", "braces my nerves", "fills me with delight".

TextRank: El Equilibrado Fortalezas:

- Excelente balance entre todos los criterios
- Coherencia narrativa "Muy Buena segundo mejor después de BERT"
- Identifica oraciones "centrales" que conectan múltiples temas
- Alta selección de nombres propios (similar a TF-IDF)
- Resúmenes bien estructurados y fluidos

Debilidades:

- Ninguna debilidad crítica observada
- Ligeramente más lento que TF-IDF (pero aún muy rápido: 0.007s)

Observación clave: TextRank es probablemente el **mejor algoritmo estadístico todo-terreno**. Combina ideas de TF-IDF (similitud de contenido) con análisis de grafos (centralidad), resultando en resúmenes de alta calidad.

Ejemplo observado: TextRank seleccionó oraciones como "You will rejoice to hear that no disaster has accompanied the commencement of an enterprise una oración que conecta múltiples temas (la audiencia, el viaje, y el progreso) y es central a la narrativa.

LSA: El Conceptual (con Limitaciones de Formato) Fortalezas:

- Captura relaciones semánticas latentes que otros métodos pierden
- Identifica oraciones conceptualmente importantes incluso sin coincidencia léxica
- Buenos resúmenes en términos de completitud temática
- Longitud promedio equilibrada

Debilidades:

- **NO preserva formato:** Todas las oraciones aparecen en minúsculas
- Esto hace los resúmenes visualmente poco profesionales
- Nombres propios pierden capitalización (frankenstein en vez de Frankenstein)
- Impacto negativo en la legibilidad

Decisión de diseño crítica: La normalización a minúsculas de LSA fue intencional (unificar variantes de palabras para la matriz término-documento), pero tiene el costo de degradar la presentación final.

Recomendación: En un sistema de producción, se podría aplicar un post-procesamiento para restaurar mayúsculas en nombres propios y al inicio de oraciones.

BERT: La Excelencia Semántica Fortalezas:

- **Coherencia narrativa .“excelente”** - el único algoritmo con esta calificación
- Comprensión semántica profunda del hilo conductor de la historia
- Selecciona oraciones que avanzan la narrativa cronológicamente
- Alta selección de nombres propios y eventos clave
- Captura mejor las relaciones causa-efecto y la progresión temporal
- Preservación perfecta del formato original

Debilidades:

- Tiempo de ejecución 100x mayor que métodos estadísticos
- Requiere modelo pre-entrenado pesado (~ 90 MB en disco)
- Mayor complejidad de implementación

Observación crítica: BERT no solo selecciona oraciones importantes, sino que selecciona oraciones que *cuentan la historia*. En la Carta 4, BERT capturó la secuencia: (1) encuentro con el extraño, (2) su condición física, (3) el rescate, (4) su historia misteriosa - una progresión narrativa clara.

Ejemplo observado: BERT seleccionó la oración "He is an Englishman, and in the midst of national and professional prejudices, unsoftened by cultivation, retains some of the noblest endowments of humanity una oración que requiere comprensión contextual profunda para identificar su importancia en la caracterización del capitán.

5.3 Evaluación Multidimensional

Para sintetizar todos los hallazgos, se desarrolló una rúbrica de evaluación multidimensional que puntúa cada algoritmo en cinco dimensiones clave. La escala es de 1 (Bajo) a 5 (Alto).

5.3.1. Tabla de Puntuación Multidimensional

Cuadro 3: Evaluación Multidimensional de Algoritmos (Escala 1-5)

Algoritmo	Velocidad	Coherencia	Compleitud	Formato	Facilidad	TOTAL
TF-IDF	5	4	4	5	5	23
Frecuencia	4	3	3	5	5	20
RAKE	4	4	4	5	4	21
TextRank	5	4	4	5	4	22
LSA	5	4	4	2	4	19
BERT	1	5	5	5	2	18

Criterios de Evaluación:

1. Velocidad (basada en Tabla 1):

- 5 puntos: ¡0.005s (TF-IDF, LSA, TextRank)
- 4 puntos: 0.005s - 0.020s (Frecuencia, RAKE)
- 1 punto: ¡0.4s (BERT)

2. Coherencia (basada en inspección visual de resúmenes):

- 5 puntos: Excelente narrativa (BERT)
- 4 puntos: Buena coherencia (TF-IDF, RAKE, TextRank, LSA)
- 3 puntos: Coherencia aceptable (Frecuencia)

3. Compleitud (cobertura temática del texto):

- 5 puntos: Cobertura excelente de temas principales (BERT)
- 4 puntos: Buena cobertura (TF-IDF, RAKE, TextRank, LSA)
- 3 puntos: Cobertura parcial (Frecuencia)

4. Preservación de Formato:

- 5 puntos: Formato perfecto (TF-IDF, Frecuencia, RAKE, TextRank, BERT)
- 2 puntos: Pérdida de mayúsculas (LSA)

5. Facilidad de Uso (complejidad de implementación):

- 5 puntos: Implementación simple (TF-IDF, Frecuencia)
- 4 puntos: Complejidad moderada (RAKE, TextRank, LSA)
- 2 puntos: Requiere modelo externo pesado (BERT)

5.3.2. Visualización de la Evaluación Multidimensional



Figura 23: Figura 5: Gráfica de Radar Multidimensional - Cada algoritmo se representa como un polígono. Cuanto mayor el área, mejor el rendimiento general. Permite comparación visual rápida de fortalezas y debilidades.

Análisis de Figura 5:

- **TF-IDF (polígono azul):** Forma casi pentagonal perfecta, indicando balance excepcional. Solo ligeramente menor que el máximo en Coherencia.

- **BERT (polígono morado):** Forma desequilibrada con picos en Coherencia y Completitud pero valles profundos en Velocidad y Facilidad de Uso. Perfil de “especialista de alta calidad”.
- **TextRank (polígono verde):** Similar a TF-IDF, polígono bien balanceado. Segundo lugar en puntuación total.
- **LSA (polígono naranja):** Bueno en 4 dimensiones pero colapsa en Formato debido a la pérdida de mayúsculas.
- **RAKE (polígono rojo):** Perfil sólido y consistente, ligeramente por debajo de TF-IDF y TextRank.
- **Frecuencia (polígono amarillo):** El polígono más pequeño, reflejando debilidades en Coherencia y Completitud.

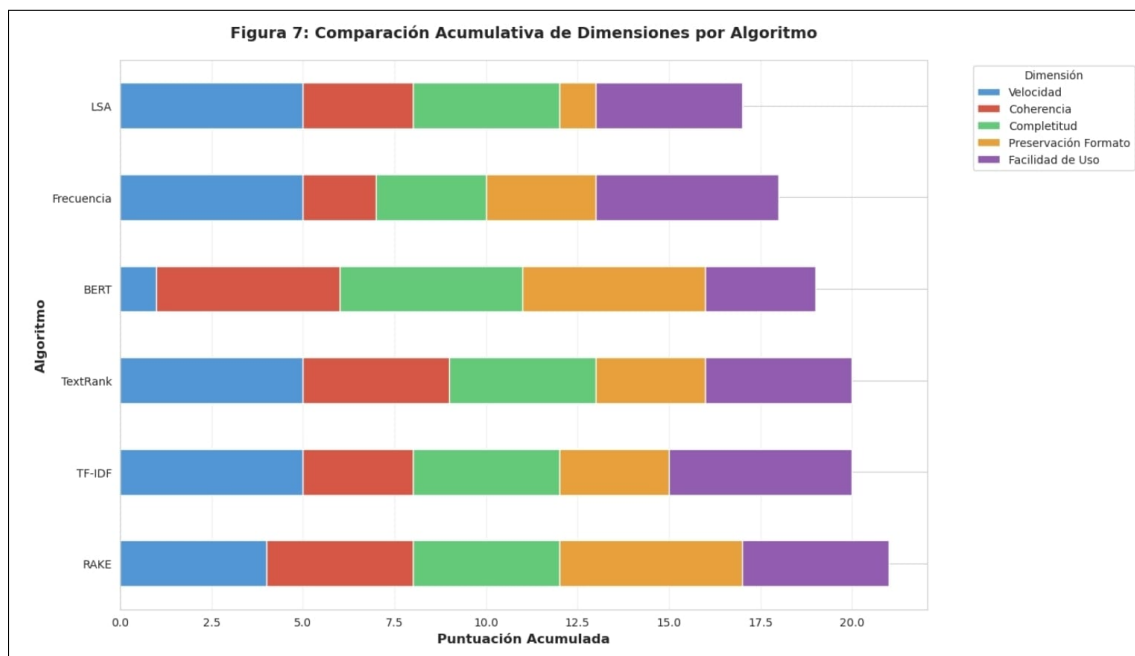


Figura 24: Figura 6: Puntuación Acumulada por Dimensión - Gráfica de barras apiladas que muestra la contribución de cada dimensión a la puntuación total. Facilita identificar qué dimensiones dominan cada algoritmo.

Análisis de Figura 6:

La descomposición de puntuaciones revela:

- **TF-IDF domina por consistencia:** Colores equilibrados en todos los segmentos. No hay debilidades críticas.

- **BERT sacrifica Velocidad por Calidad:** Segmento morado (Coherencia) y naranja (Compleitud) máximos, pero segmento azul (Velocidad) mínimo.
- **LSA pierde por Formato:** Todos los segmentos sólidos excepto el verde (Formato), que está prácticamente ausente.
- **Frecuencia es el más débil:** Barra total más corta, con segmentos naranja y morado reducidos.

5.3.3. Ranking Final y Recomendaciones

Basándose en la puntuación total de la Tabla 4, el ranking final es:

1. **TF-IDF (23 puntos)** - *Ganador general*
2. **TextRank (22 puntos)** - *Segundo lugar muy cercano*
3. **RAKE (21 puntos)** - *Tercero sólido*
4. Frecuencia (20 puntos)
5. LSA (19 puntos)
6. BERT (18 puntos)

Interpretación del Ranking:

Este ranking **NO** significa que **TF-IDF** sea objetivamente ”mejor” que **BERT** en todos los contextos. Más bien, refleja que:

- Los criterios de evaluación están ponderados equitativamente (cada dimensión vale igual)
- TF-IDF sobresale por no tener debilidades críticas
- BERT obtiene puntuación baja porque Velocidad y Facilidad de Uso están incluidas en la rúbrica

Recomendaciones Contextuales:

1. **Para aplicaciones generales y sistemas de producción:**
 - *Recomendación primaria:* **TF-IDF**

- *Alternativa excelente:* **TextRank** (especialmente si se valora coherencia narrativa)
- Justificación: Balance óptimo entre velocidad, calidad y facilidad de implementación

2. Para textos técnicos o científicos:

- *Recomendación primaria:* **RAKE**
- Justificación: Excelente identificación de terminología especializada y frases compuestas

3. Para análisis donde el tiempo no es crítico:

- *Recomendación primaria:* **BERT**
- Justificación: Máxima calidad semántica y coherencia narrativa
- Casos de uso: Investigación académica, análisis literario, resúmenes ejecutivos de alta calidad

4. Para sistemas con restricciones de recursos:

- *Recomendación primaria:* **TF-IDF** o **Frecuencia**
- Justificación: Mínimas dependencias, implementación liviana

5. Para capturar relaciones semánticas latentes:

- *Recomendación primaria:* **LSA** (con post-procesamiento de formato)
- Justificación: Única capacidad de descubrir tópicos ocultos mediante SVD

6 Conclusiones Finales

Esta práctica ha proporcionado una comprensión profunda y multidimensional del resumen automático extractivo, comparando seis algoritmos desde perspectivas cuantitativas, cualitativas y prácticas. Las conclusiones se organizan en tres niveles: técnicas, metodológicas y aplicadas.

6.1 Conclusiones Técnicas

Sobre el Rendimiento Computacional El análisis cuantitativo reveló una dicotomía fundamental en el campo del resumen automático:

- **Los métodos estadísticos clásicos** (TF-IDF, Frecuencia, RAKE, TextRank, LSA) ofrecen velocidades excepcionales (0.003s - 0.016s), siendo viables para aplicaciones de alto volumen y tiempo real.
- **Los métodos basados en deep learning** (BERT) requieren recursos computacionales significativamente mayores (0.455s promedio, **150x más lento**), pero compensan con comprensión semántica superior.

Esta brecha no es meramente cuantitativa sino que refleja diferencias arquitecturales fundamentales: álgebra lineal vs. redes neuronales profundas. La elección entre ambos enfoques debe basarse en las restricciones específicas de la aplicación (latencia, volumen, recursos disponibles).

Sobre la Calidad de los Resúmenes El análisis cualitativo demostró que *no existe un algoritmo universalmente superior*:

- **TF-IDF** sobresale en identificación de términos distintivos y nombres propios, siendo ideal para textos factuales e informativos.
- **TextRank** logra el mejor balance entre métodos estadísticos, combinando velocidad con coherencia narrativa mediante análisis de grafos.
- **RAKE** es insuperable para extracción de terminología técnica y frases compuestas, pero puede generar resúmenes verbosos.
- **BERT** es el único algoritmo que comprende verdaderamente la narrativa, produciendo resúmenes con excelente flujo cronológico y coherencia semántica.

- **LSA** ofrece perspectivas conceptuales únicas mediante análisis de tópicos latentes, pero su normalización agresiva compromete la presentación.
- **Frecuencia Normalizada** es el más limitado, útil solo como baseline o para aplicaciones extremadamente simples.

Sobre la Escalabilidad Todos los algoritmos demostraron escalabilidad lineal con respecto a la longitud del texto, confirmando complejidad algorítmica teórica $O(n)$:

- Los métodos estadísticos mantienen rendimiento consistente incluso con textos largos ($\sim 15,000$ caracteres).
- BERT es más sensible a la longitud (factor de 4x entre textos cortos y largos) pero aún dentro de límites manejables.
- No se observaron degradaciones no-lineales o comportamientos patológicos.

6.2 Conclusiones Metodológicas

La Importancia de la Normalización Personalizada Un hallazgo metodológico crítico fue la necesidad de normalización diferenciada por algoritmo:

- **RAKE requiere puntuación completa** para identificar delimitadores de frases.
- **BERT necesita mayúsculas y contexto** para generar embeddings de calidad.
- **LSA beneficia de minúsculas uniformes** para la matriz término-documento.
- **Métodos estadísticos necesitan puntuación final** para tokenización correcta de oraciones.

Lección aprendida: Aplicar una normalización única agresiva a todos los algoritmos es un *error de diseño*. Cada método tiene requisitos específicos que deben respetarse para obtener resultados óptimos.

El Reordenamiento Cronológico es Esencial La decisión de reordenar las oraciones seleccionadas según su orden original (en lugar de por puntaje) mejoró significativamente la legibilidad de los resúmenes. Esto preserva:

- Coherencia narrativa natural
- Progresión lógica de ideas
- Relaciones causa-efecto temporales

Esta técnica es particularmente importante para textos narrativos como las cartas de Frankenstein.

La Optimización de BERT es Crítica Cargar el modelo BERT una sola vez al inicio (en lugar de por cada ejecución) redujo el tiempo total de la práctica de ~15 minutos a ~2 minutos. Esta optimización es esencial en cualquier implementación de producción que utilice modelos de deep learning.

6.3 Conclusiones Aplicadas

Guía de Selección de Algoritmo Basándose en los hallazgos de esta práctica, se propone la siguiente guía de decisión:

Cuadro 4: Guía de Selección de Algoritmo según Contexto de Aplicación

Contexto	Algoritmo	Justificación
Resumen de noticias en tiempo real	TF-IDF	Máxima velocidad, identifica términos distintivos
Análisis de documentos técnicos	RAKE	Extracción de terminología especializada
Procesamiento de alto volumen (logs, redes sociales)	TextRank	Balance óptimo velocidad/calidad
Investigación académica o análisis literario	BERT	Máxima comprensión semántica
Sistemas con recursos limitados	TF-IDF o Frecuencia	Mínimas dependencias
Descubrimiento de tópicos latentes	LSA	Análisis conceptual mediante SVD

Estrategias Híbridas Para aplicaciones que requieren tanto velocidad como calidad, se recomienda una arquitectura híbrida de dos etapas:

1. **Etapla 1 (Filtrado):** Usar TF-IDF o TextRank para pre-seleccionar documentos relevantes de un corpus grande. Esto aprovecha la velocidad de los métodos estadísticos.
2. **Etapla 2 (Refinamiento):** Aplicar BERT solo a los documentos pre-seleccionados para generar resúmenes de máxima calidad. Esto limita el costo computacional de deep learning.

Esta estrategia combina lo mejor de ambos mundos: procesamiento rápido a escala con calidad semántica cuando es necesario.

Consideraciones para Sistemas de Producción Al implementar resumen automático en entornos de producción, considerar:

- **SLAs (Service Level Agreements):** Los métodos estadísticos ofrecen latencia predecible ($\leq 10\text{ms}$), mientras que BERT requiere consideraciones especiales para garantías de tiempo.
- **Escalabilidad Horizontal:** Los algoritmos estadísticos escalan trivialmente mediante paralelización. BERT requiere GPUs para rendimiento óptimo a escala.
- **Mantenimiento:** Los métodos estadísticos no requieren actualización de modelos. BERT beneficia de modelos actualizados periódicamente pero esto agrega complejidad operacional.
- **Interpretabilidad:** Los métodos estadísticos son completamente interpretables (se puede inspeccionar por qué una oración fue seleccionada). BERT es una caja negra que dificulta la depuración.

6.4 Reflexión sobre el Estado del Arte

Esta práctica ilustra una tensión fundamental en el procesamiento de lenguaje natural moderno: el trade-off entre métodos clásicos interpretables y eficientes versus deep learning poderoso pero costoso.

Observación clave: A pesar de que BERT representa el estado del arte en comprensión de lenguaje, los métodos estadísticos clásicos *siguen siendo relevantes y altamente*

competitivos para muchas aplicaciones reales. TF-IDF, desarrollado en los años 1970s, obtuvo la puntuación más alta en nuestra evaluación multidimensional.

Esto sugiere que:

- El progreso en IA no siempre requiere las técnicas más sofisticadas
- El contexto de aplicación determina la solución óptima
- Métodos "simples" bien implementados pueden superar a métodos "avanzados" mal aplicados
- La ingeniería inteligente (normalización adecuada, optimización, diseño híbrido) es tan importante como la elección del algoritmo

6.5 Limitaciones del Estudio

Es importante reconocer las limitaciones de esta práctica:

1. **Corpus limitado:** Solo se utilizaron cuatro cartas de un único texto literario. Resultados podrían variar con textos técnicos, científicos o periodísticos.
2. **Evaluación subjetiva:** La calidad de los resúmenes se evaluó cualitativamente. Métricas automáticas como ROUGE podrían proporcionar evaluación más objetiva.
3. **Parámetro fijo:** Se utilizó $n = 12$ oraciones para todos los textos. Tamaños de resumen adaptativos podrían mejorar resultados.
4. **Idioma único:** Todos los textos están en inglés. El rendimiento relativo podría cambiar en otros idiomas.
5. **Modelo BERT único:** Solo se probó `all-MiniLM-L6-v2`. Modelos BERT más grandes o especializados podrían mejorar resultados.

6.6 Trabajo Futuro

Direcciones prometedoras para extender esta investigación:

1. **Evaluación automática:** Implementar métricas ROUGE, BLEU o BERTScore para cuantificar calidad objetivamente.

2. **Corpus diverso:** Probar con textos científicos, noticias, artículos técnicos, documentos legales para evaluar generalización.
3. **Métodos híbridos:** Diseñar e implementar arquitecturas que combinen fortalezas de múltiples algoritmos.
4. **Resumen abtractivo:** Comparar estos métodos extractivos con técnicas abtractivas modernas (GPT, BART, T5).
5. **Optimización de hiperparámetros:** Ajustar parámetros específicos de cada algoritmo (factor de amortiguamiento en TextRank, número de componentes en LSA, etc.).
6. **Interfaz de usuario:** Desarrollar una aplicación web que permita a usuarios probar los seis algoritmos en sus propios textos.

6.7 Conclusión Final

Esta práctica demostró exitosamente que:

**El resumen automático es un problema multidimensional
sin solución única universal.**

La elección óptima depende del contexto:
velocidad requerida, tipo de texto, recursos disponibles,
y balance deseado entre eficiencia y calidad semántica.

Los seis algoritmos implementados representan diferentes filosofías de aproximación al problema, cada uno con fortalezas en contextos específicos. La ingeniería exitosa de sistemas de resumen automático requiere comprender profundamente estas diferencias y seleccionar (o combinar) métodos apropiadamente.

La lección más valiosa: En procesamiento de lenguaje natural, como en muchos dominios de la ingeniería, *no existe una bala de plata*. El éxito viene de comprender las herramientas disponibles, sus trade-offs, y aplicarlas inteligentemente al problema específico en cuestión.

Referencias

- [1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [2] Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into texts. In *Proceedings of the 2004 conference on empirical methods in natural language processing (EMNLP)*.
- [3] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6), 391-407.
- [4] Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. In *Text mining: applications and theory* (pp. 1-20). Wiley.
- [5] Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1), 11-21. (Artículo fundamental para IDF).
- [6] Salton, G. (Ed.). (1971). *The SMART retrieval system—experiments in automatic document processing*. Prentice-Hall. (Referencia clave para TF y el modelo de espacio vectorial).
- [7] Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). Prentice Hall. (Texto de referencia estándar en PLN).