

# RAIn 2025

## Trabajo Práctico N°4

**Tema:** Unidad 3 y 4

**Fecha Inicio:** 20/05/2025 **Fecha de Entrega:** 03/06/2025

**Autor:** Jorge Justino Riera **LU:** 5575 **Carrera:** Ing. Inf **Plan:** 2010

**Repositorio código fuente:** <https://github.com/nicrom8b/rain/tree/main/tp5>

<b>Tema: Unidad 3 y 4.....</b>	<b>1</b>
Fecha Inicio: 20/05/2025 Fecha de Entrega: 03/06/2025.....	1
Autor: Jorge Justino Riera LU: 5575 Carrera: Ing. Inf Plan: 2010.....	1
Repositorio código fuente:.....	1
<b>Actividades.....</b>	<b>2</b>
Procesador PDFs.....	3
Ejecución.....	3
Descripción.....	5
Funciones principales.....	5
Procesador de Noticias.....	7
Ejecución.....	8
Descripción.....	9
Análisis.....	9
Análisis de discurso - Resumen.....	10
Ejecución.....	10
Descripción.....	11
Funciones Principales.....	11

# Actividades

1. Para los documentos procesados en el punto N° 3 del TP N° 1 (5 documentos) y Punto N° 2 del TPN ° 3 (10 documentos/noticias), por separado, cree una representación TF-IDF de los mismos (RepA primer caso y RepB segundo caso). Las representaciones deben ser de  $n-1$  documentos, dejando 1 (un) documento para realizar una consulta y obtener el top 3 de documentos más parecidos/similares. Debe mostrar como resultado para cada representación el score obtenido y el texto del documento.

Las representaciones a realizar deben tomar:

- a. El texto original
- b. Eliminando stop-words
- c. Con bi-gramas

# Procesador PDFs

## Ejecución

```
(1-1) (* [infra:istio-ingress]jriera:1-1/ (main*) $ python main.py [23:08:48]
[nltk_data] Downloading package stopwords to
[nltk_data] /Users/jriera/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Documento de consulta: Algoritmo-vinculacion-relatos-denuncias.pdf

--- Representación: Original ---
Score: 0.8493
Archivo: Accesibilidad-web.pdf
Texto (primeros 500 chars): Accesibilidad web desde la percepción de las personas
con discapacidad: una herramienta tecnológica para su
valoración
Ivana Harari1, Sergio Luján-Mora2, Javier Díaz1, Daniel Idalgo1
1 Universidad de La Plata , La Plata BA , Argentina.
{iharari, jdiaz }info.unlp.edu.ar ; didalgo@gmail.com

2 Universidad de Alicante. Alicante, España.
sergio.lujan@ua.es
Resumen . En plena y constante evolución tecnológica, muchas personas con
discapacidad aún enfrentan barreras para acceder al con

Score: 0.8266
Archivo: Analisis-sericio-urgencias.pdf
Texto (primeros 500 chars): Análisis de desempeño de Servicio de Urgencia
Hospitalario con llegada creciente de pacientes
Mariela Rodríguez1, Francesc Boixader2, Francisco Epelde3, Eva Bruballa4, Armando
De Giusti5, Alvaro Wong6, Dolores Rexachs7 and Emilio Luque8

1 Faculty of Engineering, National University of Jujuy 4600, Argentina
2, 4 Escoles Universitaries Gimbernat (EUG), Computer Science School, Universi -tat Autònoma
de Barcelona, Sant Cugat del Vallès, Barcelona, 08174, Spain
3Consultant Internal Medicine

Score: 0.8076
Archivo: Algoritmo-aprendizaje-automatizado.pdf
Texto (primeros 500 chars): Lodato Marcos; Jaime Fernando; Lopez Pedro
CAETI - Universidad Abierta Interamericana
Av. Ovidio Lagos 944 - Rosario, Santa Fe - Argentina
{Marcos.Lodato; Fernando.Jaime }@alumnos.uai.edu.ar {Pedro.Lopez}@uai.edu.ar
Resumen. En el marco de esta investigación, se busca explorar y analizar los
conceptos claves en relación con los microcontroladores y los algoritmos de
aprendizaje automático. Se busca abordar conceptos claves sobre los
microcontroladores , especificaciones y casos de
```

--- Representación: Sin Stopwords ---

Score: 0.1780

Archivo: Accesibilidad-web.pdf

Texto (primeros 500 chars): Accesibilidad web desde la percepción de las personas con discapacidad: una herramienta tecnológica para su valoración

Ivana Harari<sup>1</sup>, Sergio Luján-Mora<sup>2</sup>, Javier Díaz<sup>1</sup>, Daniel Idalgo<sup>1</sup>

<sup>1</sup> Universidad de La Plata , La Plata BA , Argentina.

{iharari, jdiaz}@info.unlp.edu.ar ; didalgo@gmail.com

<sup>2</sup> Universidad de Alicante. Alicante, España.

sergio.lujan@ua.es

Resumen . En plena y constante evolución tecnológica, muchas personas con discapacidad aún enfrentan barreras para acceder al con

Score: 0.1690

Archivo: Algoritmo-aprendizaje-automatizado.pdf

Texto (primeros 500 chars): Lodato Marcos; Jaime Fernando; Lopez Pedro

CAETI - Universidad Abierta Interamericana

Av. Ovidio Lagos 944 - Rosario, Santa Fe - Argentina

{Marcos.Lodato; Fernando.Jaime}@alumnos.uai.edu.ar {Pedro.Lopez}@uai.edu.ar

Resumen. En el marco de esta investigación, se busca explorar y analizar los conceptos claves en relación con los microcontroladores y los algoritmos de aprendizaje automático. Se busca abordar conceptos claves sobre los microcontroladores , especificaciones y casos de

Score: 0.1625

Archivo: Analisis-sericio-urgencias.pdf

Texto (primeros 500 chars): Análisis de desempeño de Servicio de Urgencia

Hospitalario con llegada creciente de pacientes

Mariela Rodríguez<sup>1</sup>, Francesc Boixader<sup>2</sup>, Francisco Epelde<sup>3</sup>, Eva Bruballa<sup>4</sup>, Armando

De Giusti<sup>5</sup>, Alvaro Wong<sup>6</sup>, Dolores Rexachs<sup>7</sup> and Emilio Luque<sup>8</sup>

<sup>1</sup> Faculty of Engineering, National University of Jujuy 4600, Argentina

<sup>2, 4</sup> Escoles Universitaries Gimbernat (EUG), Computer Science School, Universitat Autònoma de Barcelona, Sant Cugat del Vallès, Barcelona, 08174, Spain

<sup>3</sup>Consultant Internal Medicine

```

--- Representación: Con Bigramas ---
Score: 0.8040
Archivo: Accesibilidad-web.pdf
Texto (primeros 500 chars): Accesibilidad web desde la percepción de las personas
con discapacidad: una herramienta tecnológica para su
valoración
Ivana Harari1, Sergio Luján-Mora2, Javier Díaz1, Daniel Idalgo1
1 Universidad de La Plata , La Plata BA , Argentina.
{iharari, jdiaz }info.unlp.edu.ar ; didalgo@gmail.com

2 Universidad de Alicante. Alicante, España.
sergio.lujan@ua.es
Resumen . En plena y constante evolución tecnológica, muchas personas con
discapacidad aún enfrentan barreras para acceder al con

Score: 0.7705
Archivo: Analisis-sericio-urgencias.pdf
Texto (primeros 500 chars): Análisis de desempeño de Servicio de Urgencia
Hospitalario con llegada creciente de pacientes
Mariela Rodríguez1, Francesc Boixader2, Francisco Epelde3, Eva Bruballa4, Armando
De Giusti5, Alvaro Wong6, Dolores Rexachs7 and Emilio Luque8

1 Faculty of Engineering, National University of Jujuy 4600, Argentina
2, 4 Escoles Universitaries Gimbernat (EUG), Computer Science School, Universi -tat Autònoma
de Barcelona, Sant Cugat del Vallès, Barcelona, 08174, Spain
3Consultant Internal Medici

Score: 0.7338
Archivo: Algoritmo-aprendizaje-automatizado.pdf
Texto (primeros 500 chars): Lodato Marcos; Jaime Fernando; Lopez Pedro
CAETI - Universidad Abierta Interamericana
Av. Ovidio Lagos 944 - Rosario, Santa Fe - Argentina
{Marcos.Lodato; Fernando.Jaime }@alumnos.uai.edu.ar {Pedro.Lopez}@uai.edu.ar
Resumen. En el marco de esta investigación, se busca explorar y analizar los
conceptos claves en relación con los microcontroladores y los algoritmos de
aprendizaje automático. Se busca abordar conceptos claves sobre los
microcontroladores , especificaciones y casos de

```

## Descripción

El script desarrollado busca todos los archivos PDF en una carpeta específica y extrae el texto de cada uno. Para cada texto realiza una representación numérica TF-IDF, la cual le da mas peso a las palabras importantes del texto. Esto lo realiza de tres formas o representaciones:

- Usando el texto como esta
- Quitando stop words
- Utilizando bi-gramas

Procesa todos los documentos, menos uno, el cual deja para realizar comparaciones. Estas comparaciones de similitud son realizadas usando una medida llamada “similitud de coseno”

## Funciones principales

- **extract\_text\_from\_pdf(pdf\_path)**
- **load\_documents(pdf\_dir)**
- **preprocess\_text(text, remove\_stopwords=False, use\_bigrams=False)**
- **main():** coordina todo el flujo del script: carga los textos de los archivos PDF, separa uno como documento de consulta y utiliza los demás como base de comparación. Para cada tipo de representación (texto original, sin stopwords y con bigramas), preprocesa

los textos, calcula sus vectores TF-IDF, mide la similitud entre el documento de consulta y los del corpus, y finalmente muestra en pantalla los tres documentos más similares junto con su puntaje y un fragmento de su contenido.

```
def main():
    docs = load_documents(PDF_DIR)
    if len(docs) < 2:
        print('Se requieren al menos 2 documentos PDF.')
        return
    # Dejar uno para consulta
    query_doc = docs[-1]
    corpus_docs = docs[:-1]

    print(f"Documento de consulta: {query_doc['filename']}")
    print()

    for rep in REPRESENTATIONS:
        print(f"--- Representación: {rep['name']} ---")
        # Preprocesar
        corpus = [preprocess_text(d['text'], rep['remove_stopwords'],
rep['use_bigrams']) for d in corpus_docs]
        query = preprocess_text(query_doc['text'], rep['remove_stopwords'],
rep['use_bigrams'])
        # TF-IDF
        vectorizer = TfidfVectorizer()
        X = vectorizer.fit_transform(corpus)
        q_vec = vectorizer.transform([query])
        # Similitud
        scores = cosine_similarity(q_vec, X)[0]
        # Top 3
        top_idx = scores.argsort()[::-1][:3]
        for idx in top_idx:
            print(f"Score: {scores[idx]:.4f}")
            print(f"Archivo: {corpus_docs[idx]['filename']}")
            print(f"Texto (primeros 500 chars): {corpus_docs[idx]['text'][:500]}")
            print()
        print()
```

# Procesador de Noticias

## Ejecución

```
(1-2) (* |infra:istio-ingress)jriera:1-2/ (mainx) $ python main.py [23:35:21]
[nltk_data] Downloading package stopwords to
[nltk_data] /Users/jriera/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Noticia de consulta: El tributarista César Litvin explicó las nuevas medidas del Gobierno: "Esto no significa que a futuro sea un 'viva la pepa'"

--- Representación: Original (TF-IDF RepB) ---
Score: 0.6792
Título: Cómo se va a poder usar la plata del colchón para pagos en el supermercado, compra de autos y otros fines
Texto (primeros 500 chars): 23 May, 2025
PorAgustín Maza
Las recientesmedidas del Gobierno deJavier Mileiabrieron la puerta para que los argentinos que tienen ahorros sin declarar, como los que poseen "dólares debajo del colchón", los utilicen para realizar toda clase de operaciones como compra de autos, departamento s, inversiones y consumos personales sin ser alcanzados por el radar de la Agencia de Recaudación y Control Aduanero (ARCA). Sin embargo, los tributaristas advierten que hará falta un cambio en la legislación para gen

Score: 0.6686
Título: Luz y gas: tras los aumentos, las empresas invertirán más de USD 5.900 millones para mejorar el servicio y reducir cortes
Texto (primeros 500 chars): 23 May, 2025
Las empresas de distribución y transporte de gas natural y energía eléctrica comprometieron ante el Gobierno inversiones por más de USD 5.900 millones en los próximos cinco años a cambio de mejoras en el servicio, según informó este viernes la Secretaría de Energía de la Nación.
El compromiso se dio en el marco de la Revisión Quinquenal de Tarifas (RQT), en donde las compañías (Edesur, Edenor, Metrogas, Camuzzi, Naturgy, TGS, TGN, entre otras) comprometieron desembolsos a cambio de los incrementos

Score: 0.6641
Título: Dólar hoy en vivo: bajó a $1.150 en el Banco Nación y subió a $1.170 en el mercado blue
Texto (primeros 500 chars): 23 May, 2025
Las reservas internacionales brutas del Banco Central se incrementaron este viernes en USD 329 millones o 0,9%, a USD 38.425 millones. En la última semana este stock de activos creció en 93 millones de dólares.
Fuentes del BCRA atribuyeron el aumento "principalmente, a la subida de cotizaciones de monedas y por un desembolso del BID (Banco Interamericano de Desarrollo) de USD 135 millones". En Wall Street el oro ganó 2,1% y el yuan chino, un 0,3 por ciento.
Los contratos de dólar futuro
```

```
--- Representación: Sin Stopwords (TF-IDF RepB) ---
Score: 0.3091
Título: Cómo se va a poder usar la plata del colchón para pagos en el supermercado, compra de autos y otros fines
Texto (primeros 500 chars): 23 May, 2025
PorAgustín Maza
Las recientesmedidas del Gobierno deJavier Mileiabrieron la puerta para que los argentinos que tienen ahorros sin declarar, como los que poseen "dólares debajo del colchón", los utilicen para realizar toda clase de operaciones como compra de autos, departamento s, inversiones y consumos personales sin ser alcanzados por el radar de la Agencia de Recaudación y Control Aduanero (ARCA). Sin embargo, los tributaristas advierten que hará falta un cambio en la legislación para gen

Score: 0.1398
Título: Dólar hoy en vivo: bajó a $1.150 en el Banco Nación y subió a $1.170 en el mercado blue
Texto (primeros 500 chars): 23 May, 2025
Las reservas internacionales brutas del Banco Central se incrementaron este viernes en USD 329 millones o 0,9%, a USD 38.425 millones. En la última semana este stock de activos creció en 93 millones de dólares.
Fuentes del BCRA atribuyeron el aumento "principalmente, a la subida de cotizaciones de monedas y por un desembolso del BID (Banco Interamericano de Desarrollo) de USD 135 millones". En Wall Street el oro ganó 2,1% y el yuan chino, un 0,3 por ciento.
Los contratos de dólar futuro

Score: 0.1287
Título: Las acciones y los bonos argentinos resistieron con solidez otra rueda negativa de Wall Street
Texto (primeros 500 chars): 23 May, 2025
PorJuan Gasalla
La plaza financiera argentina atravesó con saldo positivo una ronda de negocios amenazada por la tendencia bajista de los mercados internacionales.
Las acciones mundiales y el dólar cayeron este viernes después de que el presidente de Estados Unidos, Donald Trump, lanzó su más reciente "bomba" comercial al recomendar aranceles del 50% sobre las importaciones de la Unión Europea a partir del 1 de junio. Además, el mandatario amenazó a Apple (-3%) con un aumento de aranceles
```

--- Representación: Con Bigramas (TF-IDF RepB) ---

Score: 0.5504

Título: Cómo se va a poder usar la plata del colchón para pagos en el supermercado, compra de autos y otros fines

Texto (primeros 500 chars): 23 May, 2025

Por Agustín Maza

Las recientes medidas del Gobierno de Javier Milei abrieron la puerta para que los argentinos que tienen ahorros sin declarar, como los que poseen "dólares debajo del colchón", los utilicen para realizar toda clase de operaciones como compra de autos, departamento, inversiones y consumos personales sin ser alcanzados por el radar de la Agencia de Recaudación y Control Aduanero (ARCA). Sin embargo, los tributaristas advierten que hará falta un cambio en la legislación para generar

Score: 0.5321

Título: Habló la argentina que soñaba con tener una empresa "unicornio" y terminó denunciada en EEUU por un supuesto fraude de USD 100 millones

Texto (primeros 500 chars): 23 May, 2025

La argentina Silvina Moschini fue acusada por la Comisión de Bolsa y Valores (SEC) por una presunta defraudación a los inversores de la empresa de la que es CEO. El fraude, dijo el regulador estadounidense, alcanza los USD 100 millones. Moschini es fundadora y CEO de Unicoi Inc., una empresa con sede en Nueva York, a la que definió como "America First crypto company", o la primera empresa cripto de EEUU. También fueron denunciados Alex Konanykhin, un empresario ruso que es el actual pr

Score: 0.5286

Título: Luz y gas: tras los aumentos, las empresas invertirán más de USD 5.900 millones para mejorar el servicio y reducir cortes

Texto (primeros 500 chars): 23 May, 2025

Las empresas de distribución y transporte de gas natural y energía eléctrica comprometieron ante el Gobierno inversiones por más de USD 5.900 millones en los próximos cinco años a cambio de mejoras en el servicio, según informó este viernes la Secretaría de Energía de la Nación.

El compromiso se dio en el marco de la Revisión Quinquenal de Tarifas (RQT), en donde las compañías (Edesur, Edenor, Metrogas, Camuzzi, Naturgy, TGS, TGN, entre otras) comprometieron desembolsos a cambio de los incrementos

## Descripción

El proyecto carga las 10 primeras noticias de un archivo JSON, separa una como noticia de consulta y utiliza las otras como base de comparación. Para cada una de tres representaciones del texto (original, sin stopwords y con bigramas), transforma los textos en vectores numéricos usando TF-IDF, calcula la similitud entre la noticia de consulta y las demás, y muestra en pantalla los tres artículos más similares junto con su puntaje y un fragmento de su contenido.

## Análisis

La forma en que presentamos un texto tiene un impacto directo en los resultados de comparación. Cambios en el procesamiento, como mantener el texto original, quitar palabras comunes o agregar pares de palabras, pueden modificar qué documentos parecen más similares y en qué medida lo son. Esta adaptabilidad es útil porque permite personalizar el análisis según lo que busquemos: ya sea encontrar coincidencias exactas de palabras, identificar los temas principales o destacar frases importantes.

2. Empleando el corpus inaugural de NLTK (que contiene los discursos inaugurales de los presidentes de EEUU), utilice el archivo '2009-Obama.txt' y realice lo siguiente. Obtenga un resumen del texto entre 7 y 10 oraciones, mediante 2 métodos distintos de resúmenes extractivos, compare los resultados.



# Análisis de discurso - Resumen

## Ejecución

```
(2) (* |infra:istio-ingress)jriera:2/ (main*) $ python main.py [23:45:20]
[nltk_data] Downloading package inaugural to
[nltk_data] /Users/jriera/nltk_data...
[nltk_data] Package inaugural is already up-to-date!
[nltk_data] Downloading package punkt to /Users/jriera/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] /Users/jriera/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
--- Resumen por Frecuencia de Palabras ---
- We are shaped by every language and culture, drawn from every end of this Earth; and because we have tasted the bitter swill o
f civil war and segregation, and emerged from that dark chapter stronger and more united, we cannot help but believe that the ol
d hatreds shall someday pass; that the lines of tribe shall soon dissolve; that as the world grows smaller, our common humanity
shall reveal itself; and that America must play its role in ushering in a new era of peace.
- Nor is the question before us whether the market is a force for good or ill. Its power to generate wealth and expand freedom i
s unmatched, but this crisis has reminded us that without a watchful eye, the market can spin out of control — the nation canno
t prosper long when it favors only the prosperous.
- What is required of us now is a new era of responsibility — a recognition, on the part of every American, that we have duties
to ourselves, our nation, and the world, duties that we do not grudgingly accept but rather seize gladly, firm in the knowledge
that there is nothing so satisfying to the spirit, so defining of our character, than giving our all to a difficult task.
- Let it be said by our children's children that when we were tested we refused to let this journey end, that we did not turn ba
ck nor did we falter; and with eyes fixed on the horizon and God's grace upon us, we carried forth that great gift of freedom an
d delivered it safely to future generations.
- At a moment when the outcome of our revolution was most in doubt, the father of our nation ordered these words be read to the
people:

"Let it be told to the future world ... that in the depth of winter, when nothing but hope and virtue could survive ... that the
city and the country, alarmed at one common danger, came forth to meet ...
- And so to all the other peoples and governments who are watching today, from the grandest capitals to the small village where
my father was born: know that America is a friend of each nation and every man, woman, and child who seeks a future of peace and
dignity, and we are ready to lead once more.
- This is the meaning of our liberty and our creed — why men and women and children of every race and every faith can join in c
elebration across this magnificent mall, and why a man whose father less than sixty years ago might not have been served at a lo
cal restaurant can now stand before you to take a most sacred oath.
- Rather, it has been the risk-takers, the doers, the makers of things'some celebrated but more often men and women obscure in t
heir labor, who have carried us up the long, rugged path towards prosperity and freedom.
```

```
--- Resumen por TextRank ---
- At a moment when the outcome of our revolution was most in doubt, the father of our nation ordered these words be read to the
people:

"Let it be told to the future world ... that in the depth of winter, when nothing but hope and virtue could survive ... that the
city and the country, alarmed at one common danger, came forth to meet ...
- But know this, America — they will be met.
- The time has come to reaffirm our enduring spirit; to choose our better history; to carry forward that precious gift, that nob
le idea, passed on from generation to generation: the God-given promise that all are equal, all are free, and all deserve a chan
ce to pursue their full measure of happiness.
- America!
- And so to all the other peoples and governments who are watching today, from the grandest capitals to the small village where
my father was born: know that America is a friend of each nation and every man, woman, and child who seeks a future of peace and
dignity, and we are ready to lead once more.
- And God bless the United States of America.
- We remain a young nation, but in the words of Scripture, the time has come to set aside childish things.
- Starting today, we must pick ourselves up, dust ourselves off, and begin again the work of remaking America.

--- Comparación ---
Oraciones en común: 2
Las oraciones en común son:
- At a moment when the outcome of our revolution was most in doubt, the father of our nation ordered these words be read to the
people:

"Let it be told to the future world ... that in the depth of winter, when nothing but hope and virtue could survive ... that the
city and the country, alarmed at one common danger, came forth to meet ...
- And so to all the other peoples and governments who are watching today, from the grandest capitals to the small village where
my father was born: know that America is a friend of each nation and every man, woman, and child who seeks a future of peace and
dignity, and we are ready to lead once more.
```

## Descripción

El proyecto toma el discurso inaugural de Obama de 2009 (usando el corpus de NLTK) y genera dos resúmenes automáticos de entre 7 y 10 oraciones: uno seleccionando las frases más importantes según la frecuencia de palabras clave, y otro usando un método de grafos (TextRank) que elige las oraciones más relevantes por su relación con el resto del texto. Al final, muestra ambos resúmenes y te dice cuántas oraciones coinciden entre los dos métodos, permitiéndote comparar fácilmente los resultados de ambos enfoques extractivos.

**Nota:** Puedes ajustar la cantidad de oraciones cambiando el parámetro `n\_sentences` en el código

## Funciones Principales

- **get\_obama\_text()**
- **summarize\_frequency(text, n\_sentences=8):** Analiza el texto, cuenta qué palabras importantes aparecen más seguido y selecciona las oraciones que contienen más de esas palabras frecuentes.
- **summarize\_textrank(text, n\_sentences=8):** Divide el texto en oraciones, calcula qué tan parecidas son entre sí usando TF-IDF, construye un grafo de relaciones y usa el algoritmo PageRank para elegir las oraciones más “conectadas” o relevantes.
- **main():** Carga el texto, genera ambos resúmenes, los muestra en pantalla y compara cuántas oraciones coinciden entre los dos métodos.

```
import nltk
import numpy as np
import networkx as nx
from nltk.corpus import inaugural, stopwords
from nltk.tokenize import sent_tokenize, word_tokenize
from sklearn.feature_extraction.text import TfidfVectorizer

nltk.download('inaugural')
nltk.download('punkt')
nltk.download('stopwords')

# Cargar el texto de Obama 2009
def get_obama_text():
    raw = inaugural.raw('2009-Obama.txt')
    return raw

# Método 1: Resumen por frecuencia de palabras

def summarize_frequency(text, n_sentences=8):
```

```

stop_words = set(stopwords.words('english'))
sentences = sent_tokenize(text)
words = word_tokenize(text.lower())
words = [w for w in words if w.isalnum() and w not in stop_words]
freq = nltk.FreqDist(words)
sentence_scores = {}
for sent in sentences:
    sent_words = word_tokenize(sent.lower())
    sent_words = [w for w in sent_words if w.isalnum()]
    score = sum(freq[w] for w in sent_words if w in freq)
    sentence_scores[sent] = score
ranked = sorted(sentence_scores, key=sentence_scores.get, reverse=True)
summary = ranked[:n_sentences]
return summary

# Método 2: TextRank (grafo de similitud de oraciones)
def summarize_textrank(text, n_sentences=8):
    sentences = sent_tokenize(text)
    vectorizer = TfidfVectorizer(stop_words='english')
    X = vectorizer.fit_transform(sentences)
    sim_matrix = (X * X.T).toarray()
    np.fill_diagonal(sim_matrix, 0)
    nx_graph = nx.from_numpy_array(sim_matrix)
    scores = nx.pagerank(nx_graph)
    ranked = sorted((scores[i], s) for i, s in enumerate(sentences)),
reverse=True)
    summary = [s for _, s in ranked[:n_sentences]]
    return summary

def main():
    text = get_obama_text()
    print('--- Resumen por Frecuencia de Palabras ---')
    freq_summary = summarize_frequency(text, n_sentences=8)
    for s in freq_summary:
        print('-', s)
    print('\n--- Resumen por TextRank ---')
    textrank_summary = summarize_textrank(text, n_sentences=8)
    for s in textrank_summary:
        print('-', s)
    print('\n--- Comparación ---')
    print(f'Oraciones      en      común:      {len(set(freq_summary)
set(textrank_summary))}')

```

```
print('Las oraciones en común son:')  
for s in set(freq_summary) & set(textrank_summary):  
    print('-', s)  
  
if __name__ == '__main__':  
    main()
```