

# Copia\_de\_ResumenTfIdf

May 21, 2024

## 1 Actividad Generar Estrategias para Generar Resúmenes

### 1.1 Grupo 6BM1

- Integrantes:
- Carmona Serrano Ian Carlo
- Rosales Benhumea Aldo

```
[27]: import pandas as pd
import numpy as np
import textwrap
import nltk
from nltk.corpus import stopwords
from nltk import word_tokenize
from sklearn.feature_extraction.text import TfidfVectorizer
from nltk.tokenize import sent_tokenize, word_tokenize
from nltk.probability import FreqDist
```

```
[28]: nltk.download('punkt')
nltk.download('stopwords')
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\ianca\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\ianca\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

```
[28]: True
```

```
[29]: df = pd.read_csv('./news.csv')
```

Extraemos sólo un documento para trabajar con él

```
[30]: doc = df['news'].sample()
print(doc.iloc[0])
```

El nuevo programa de Avión Solidario de Latam en alianza con la Fundación Cardioinfantil impulsa la atención integral a niños y niñas que padecen enfermedades cardiovasculares en los diferentes territorios del país de escasos

recursos y que requieren atención médica especializada para el manejo de sus patologías. Esta alianza se realiza bajo la dirección del programa Regale una vida. El programa se ofrece a prestar el traslado gratuito e inmediato del paciente personal médico y familiares para su cuidado y seguimiento del proceso hacia los centros poblados donde se prestará la atención especializada. La conectividad para nosotros es fundamental para poder cumplir con nuestra labor. Una vez finalizan las brigadas de diagnóstico cardiovascular los niños y sus familias deben ser trasladados a LaCardio nuestro hospital ubicado en Bogotá y es ahí donde entran aliados estratégicos como Latam para poder traer a estos pacientes a sus tratamientos y regresarlos a sus ciudades de origen comentó el Dr. Juan Gabriel Cendales director ejecutivo de la Fundación Cardioinfantil LaCardio. La problemática que busca abordar esta nueva alianza está centrada en la falta de cobertura para el diagnóstico atención y tratamiento de niños con afectaciones cardiovasculares en las periferias y poblaciones rurales del territorio. Siguiendo la hoja de ruta de atención integral a través de la red de destinos de LATAM el equipo de la Fundación Cardioinfantil realizará brigadas de diagnóstico y desplegará una serie de oportunidades de acceso a atención médica de especializada a los pacientes de Casanare Atlántico Eje Cafetero Córdoba Huila Bolívar Norte de Santander Cesar entre otras. A la fecha el programa Regale una vida cuenta con una trayectoria de atención de más de 25 años en el sector con sus brigadas de amplio alcance hacia las comunidades ha beneficiado a 54.756 niños y ha atendido integralmente a 16.495 niños y niñas.

```
[31]: doc = textwrap.fill(doc.iloc[0], replace_whitespace=False,
    ↪fix_sentence_endings=True)
    print(doc)
```

El nuevo programa de Avión Solidario de Latam en alianza con la Fundación Cardioinfantil impulsa la atención integral a niños y niñas que padecen enfermedades cardiovasculares en los diferentes territorios del país de escasos recursos y que requieren atención médica especializada para el manejo de sus patologías. Esta alianza se realiza bajo la dirección del programa Regale una vida. El programa se ofrece a prestar el traslado gratuito e inmediato del paciente personal médico y familiares para su cuidado y seguimiento del proceso hacia los centros poblados donde se prestará la atención especializada. La conectividad para nosotros es fundamental para poder cumplir con nuestra labor. Una vez finalizan las brigadas de diagnóstico cardiovascular los niños y sus familias deben ser trasladados a LaCardio nuestro hospital ubicado en Bogotá y es ahí donde entran aliados estratégicos como Latam para poder traer a estos pacientes a sus tratamientos y regresarlos a sus ciudades de origen comentó el Dr. Juan Gabriel Cendales director ejecutivo de la Fundación Cardioinfantil LaCardio. La problemática que busca abordar esta nueva alianza está centrada en la falta de cobertura para el diagnóstico atención y tratamiento de niños con afectaciones cardiovasculares en las periferias y poblaciones rurales del territorio. Siguiendo la hoja de ruta de atención integral a través

de la red de destinos de LATAM el equipo de la Fundación Cardioinfantil realizará brigadas de diagnóstico y desplegará una serie de oportunidades de acceso a atención médica de especializada a los pacientes de Casanare Atlántico Eje Cafetero Córdoba Huila Bolívar Norte de Santander Cesar entre otras. A la fecha el programa Regale una vida cuenta con una trayectoria de atención de más de 25 años en el sector con sus brigadas de amplio alcance hacia las comunidades ha beneficiado a 54.756 niños y ha atendido integralmente a 16.495 niños y niñas.

Dividir en frases

```
[32]: lineas = doc.split(".")
      lineas
```

```
[32]: ['El nuevo programa de Avión Solidario de Latam en alianza con la\nFundación
Cardioinfantil impulsa la atención integral a niños y niñas\nque padecen
enfermedades cardiovasculares en los diferentes\nterritorios del país de escasos
recursos y que requieren atención\nmédica especializada para el manejo de sus
patologías',
      ' Esta alianza\nse realiza bajo la dirección del programa Regale una vida',
      'El programa\nse ofrece a prestar el traslado gratuito e inmediato del
paciente\npersonal médico y familiares para su cuidado y seguimiento del
proceso\nhacia los centros poblados donde se prestará la
atención\nespecializada',
      'La conectividad para nosotros es fundamental para poder\ncumplir con nuestra
labor',
      ' Una vez finalizan las brigadas de\ndiagnóstico cardiovascular los niños y
sus familias deben ser\ntrasladados a LaCardio nuestro hospital ubicado en
Bogotá y es ahí\ndonde entran aliados estratégicos como Latam para poder traer a
estos\npacientes a sus tratamientos y regresarlos a sus ciudades de
origen\ncomentó el Dr',
      ' Juan Gabriel Cendales director ejecutivo de la\nFundación
CardioinfantilLaCardio',
      'La problemática que busca abordar\nesta nueva alianza está centrada en la
falta de cobertura para el\ndiagnóstico atención y tratamiento de niños con
afectaciones\ncardiovasculares en las periferias y poblaciones rurales
del\nterritorio',
      ' Siguiendo la hoja de ruta de atención integral a través\nde la red de
destinos de LATAM el equipo de la Fundación\nCardioinfantil realizará brigadas
de diagnóstico y desplegará una\nserie de oportunidades de acceso a atención
médica de especializada a\nlos pacientes de Casanare Atlántico Eje Cafetero
Córdoba Huila Bolívar\nNorte de Santander Cesar entre otras',
      'A la fecha el programa Regale una\nvida cuenta con una trayectoria de atención
de más de 25 años en el\nsector con sus brigadas de amplio alcance hacia las
comunidades ha\nbeneficiado a 54',
      '756 niños y ha atendido integralmente a 16',
      '495 niños\ny niñas',
```

```
'']
```

```
[33]: tokens = TfidfVectorizer(  
      stop_words=stopwords.words('spanish'),  
      norm='l1',  
      )  
X = tokens.fit_transform(lineas)  
X
```

```
[33]: <12x134 sparse matrix of type '<class 'numpy.float64'>'  
      with 167 stored elements in Compressed Sparse Row format>
```

```
[34]: filas, columnas = X.shape  
  
for i in range(filas):  
    for j in range(columnas):  
        print(X[i, j], end=' ')  
    print()
```

```
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.031792828748897585 0.0  
0.05175185841773685 0.0 0.0 0.041907835322369656 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0359909357823405 0.0 0.0 0.0359909357823405 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.041907835322369656 0.0 0.0 0.0 0.0 0.0  
0.041907835322369656 0.0 0.0 0.041907835322369656 0.031792828748897585 0.0 0.0  
0.0 0.0 0.0 0.0 0.031792828748897585 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.041907835322369656 0.0 0.0359909357823405 0.0 0.0 0.0 0.0 0.031792828748897585  
0.041907835322369656 0.0359909357823405 0.0 0.0359909357823405  
0.025875929208868425 0.0 0.0 0.041907835322369656 0.0 0.0 0.0 0.0 0.0  
0.041907835322369656 0.041907835322369656 0.041907835322369656 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.02853652395292657 0.0 0.0 0.041907835322369656 0.0 0.0 0.0  
0.041907835322369656 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.041907835322369656 0.0  
0.041907835322369656 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.12321143250799478 0.0 0.0 0.0 0.0  
0.0 0.0 0.16241160747790814 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.16241160747790814 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.11059179486067927 0.16241160747790814 0.0 0.0 0.0 0.13948097509880075 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.13948097509880075  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.03446384578604632 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.055816552987096896 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.055816552987096896 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.04234449469465457 0.0  
0.0 0.055816552987096896 0.0 0.0 0.0 0.0 0.0 0.0 0.055816552987096896  
0.04793590407848865 0.0 0.0 0.0 0.0 0.055816552987096896 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.055816552987096896 0.0 0.0 0.0 0.0 0.055816552987096896 0.0 0.0
```

0.055816552987096896 0.0 0.0 0.0 0.0 0.0 0.055816552987096896 0.0  
0.055816552987096896 0.0 0.055816552987096896 0.055816552987096896 0.0  
0.055816552987096896 0.038007460634357115 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.055816552987096896 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.055816552987096896 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.20581164264059598 0.0 0.0 0.20581164264059598 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.20581164264059598  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.20581164264059598 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.17675342943761613 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.037578469994873354 0.0 0.037578469994873354  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.037578469994873354 0.0  
0.028508412615501317 0.0 0.0 0.0 0.0 0.037578469994873354 0.0 0.0 0.0 0.0 0.0  
0.0 0.037578469994873354 0.0 0.037578469994873354 0.0 0.0 0.0 0.0 0.0  
0.037578469994873354 0.0 0.0 0.028508412615501317 0.0 0.0 0.0  
0.037578469994873354 0.0 0.0 0.0 0.037578469994873354 0.0 0.0 0.0  
0.037578469994873354 0.0 0.0 0.037578469994873354 0.0 0.037578469994873354 0.0  
0.0 0.0 0.0 0.0 0.0 0.037578469994873354 0.0 0.0 0.0 0.0 0.0 0.0  
0.037578469994873354 0.028508412615501317 0.0 0.0 0.0 0.0 0.02320276916917947  
0.0 0.0 0.0 0.0 0.0 0.037578469994873354 0.0 0.032272826548551514 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.032272826548551514 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.037578469994873354 0.0 0.0 0.0 0.0 0.0 0.037578469994873354 0.0 0.0 0.0  
0.0 0.0 0.037578469994873354 0.037578469994873354 0.0 0.0 0.037578469994873354  
0.0 0.0 0.037578469994873354 0.037578469994873354 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.14795882959578244 0.0 0.0 0.0 0.14795882959578244 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.14795882959578244 0.0 0.0 0.14795882959578244 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.11224702242530536 0.0 0.14795882959578244 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.14795882959578244 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.06020113571677437 0.0 0.06020113571677437 0.0 0.0 0.0  
0.045670800784857274 0.0 0.03717109972676707 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.06020113571677437 0.0 0.0 0.0 0.0 0.051701434658684174 0.0 0.0  
0.06020113571677437 0.0 0.0 0.0 0.06020113571677437 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.045670800784857274 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.06020113571677437 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.03717109972676707 0.0 0.06020113571677437  
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.06020113571677437 0.0 0.06020113571677437  
0.0 0.0 0.0 0.0 0.06020113571677437 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
0.06020113571677437 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.06020113571677437 0.0 0.0  
0.0 0.0 0.06020113571677437 0.0 0.0 0.0 0.0 0.0  
0.0 0.0 0.0 0.0 0.0 0.0 0.032826354640856155 0.0 0.0 0.0 0.0 0.0 0.0

0.040537165536612775 0.0 0.032826354640856155 0.0 0.0 0.0 0.0 0.0  
 0.032826354640856155 0.024903282727901755 0.0 0.032826354640856155  
 0.028191654681260794 0.0 0.0 0.0 0.032826354640856155 0.0 0.0 0.0  
 0.032826354640856155 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.032826354640856155 0.0  
 0.032826354640856155 0.032826354640856155 0.024903282727901755 0.0 0.0 0.0 0.0  
 0.032826354640856155 0.0 0.0 0.0 0.032826354640856155 0.0 0.024903282727901755  
 0.0 0.0 0.0 0.0 0.0 0.0 0.024903282727901755 0.0 0.0 0.0 0.0  
 0.032826354640856155 0.0 0.032826354640856155 0.0 0.0 0.028191654681260794 0.0  
 0.0 0.0 0.0 0.024903282727901755 0.0 0.028191654681260794 0.0 0.0 0.0  
 0.032826354640856155 0.0 0.0 0.0 0.032826354640856155 0.0 0.0  
 0.028191654681260794 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.032826354640856155 0.0 0.032826354640856155 0.0 0.0 0.0 0.0  
 0.032826354640856155 0.032826354640856155 0.0 0.0 0.0 0.032826354640856155  
 0.032826354640856155 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.032826354640856155 0.0  
 0.0 0.0 0.0  
 0.0 0.06396538508303855 0.0 0.06396538508303855 0.0 0.0 0.0 0.0 0.0  
 0.06396538508303855 0.0 0.0 0.06396538508303855 0.03949532977531809 0.0 0.0 0.0  
 0.06396538508303855 0.0 0.06396538508303855 0.0 0.0 0.04852649911785855 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.06396538508303855 0.0  
 0.06396538508303855 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.06396538508303855 0.0 0.0 0.0 0.0  
 0.054934215740498095 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.04355628797190491 0.0 0.0 0.0 0.0 0.054934215740498095 0.0 0.0 0.0 0.0 0.0  
 0.06396538508303855 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.06396538508303855 0.0 0.0 0.054934215740498095  
 0.2165698230068706 0.0 0.0 0.0 0.2165698230068706 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.2165698230068706 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.2165698230068706 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.13372070797251753 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.403834799770423 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.3468180170464073 0.24934718318316978 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0  
 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

```
[35]: def obtener_score(tfidf_row):  
        x = tfidf_row[tfidf_row != 0]  
        return x.mean()  
scores = np.zeros(len(lineas))  
for i in range(len(lineas)):  
    score = obtener_score(X[i,:])  
    scores[i] = score
```

c:\Python311\Lib\site-packages\scipy\sparse\\_base.py:651: RuntimeWarning: divide by zero encountered in divide

```
    return self.astype(np.float_)._mul_scalar(1./other)
```

```
[36]: scores  
sort_idx = np.argsort(-scores)  
sort_idx
```

```
[36]: array([10,  3,  9,  1,  5,  8,  6,  2,  0,  4,  7, 11], dtype=int64)
```

```
[37]: print("Resumen:")  
oraciones = []  
for i in range(0,5):  
    oraciones.append([sort_idx[i], scores[sort_idx[i]], lineas[sort_idx[i]]])  
    print(f"{scores[sort_idx[i]]}: {lineas[sort_idx[i]]}")
```

Resumen:

0.3333333333333333: 495 niños

y niñas

0.2: La conectividad para nosotros es fundamental para poder cumplir con nuestra labor

0.2: 756 niños y ha atendido integralmente a 16

0.14285714285714285: Esta alianza

se realiza bajo la dirección del programa Regale una vida

0.14285714285714285: Juan Gabriel Cendales director ejecutivo de la Fundación CardioinfantillLaCardio

Ordenar la lista por el primer elemento de cada sublista

```
[38]: oraciones_ord = sorted(oraciones, key=lambda x: x[0])  
  
for item in oraciones_ord:  
    print(item[2])
```

Esta alianza

se realiza bajo la dirección del programa Regale una vida

La conectividad para nosotros es fundamental para poder

cumplir con nuestra labor

Juan Gabriel Cendales director ejecutivo de la Fundación CardioinfantillLaCardio

756 niños y ha atendido integralmente a 16  
495 niños  
y niñas

## 1.2 Ahora vamos a probar otra tecnica para obtener palabras relevantes

### 1.2.1 Pruebas de este enfoque el cual de las palabras relevantes resaliza un resurme

```
[39]: nltk.download('punkt')
      nltk.download('stopwords')

# Lista de textos de ejemplo en español
textos = [
    'diario oficial miercoles 8 de mayo de 2024 miercoles 8 de mayo de 2024',
    'diario oficial banco de mexico tipo de cambio para solventar obligaciones',
    'denominadas en moneda extranjera pagaderas en la republica mexicana. al',
    'margen un logotipo, que dice: banco de mexico.- ''2024, ano de felipe',
    'carrillo puerto, benemerito del proletariado, revolucionario y defensor del',
    'mayab''. tipo de cambio para solventar obligaciones denominadas en moneda',
    'extranjera pagaderas en la republica mexicana el banco de mexico, con',
    'fundamento en los articulos 8o. de la ley monetaria de los estados unidos',
    'mexicanos; 35 de la ley del banco de mexico, asi como 8o. y 10 del',
    'reglamento interior del banco de mexico, y segun lo previsto en el capitulo',
    'v del titulo tercero de su circular 3/2012, informa que el tipo de cambio',
    'obtenido el dia de hoy fue de $16.9083 m.n. (dieciseis pesos con nueve mil',
    'ochenta y tres diezmilesimos moneda nacional) por un dolar de los ee.uu.a.',
    'la equivalencia del peso mexicano con otras monedas extranjeras se calculara',
    'atendiendo a la cotizacion que rija para estas ultimas contra el dolar de',
    'los ee.uu.a., en los mercados internacionales el dia en que se haga el pago.',
    'estas cotizaciones seran dadas a conocer, a solicitud de los interesados,',
    'por las instituciones de credito del pais. atentamente, ciudad de mexico, a',
    '7 de mayo de 2024.- banco de mexico: directora de disposiciones de banca',
    'central, lic. maria teresa munoz aramburu.- rubrica.- gerente de analisis de',
    'mercados nacionales, lic. dafne ramos ruiz.- rubrica. tasas de interes',
    'interbancarias de equilibrio. al margen un logotipo, que dice: banco de',
    'mexico.- ''2024, ano de felipe carrillo puerto, benemerito del proletariado,',
    'revolucionario y defensor del mayab''. tasas de interes interbancarias de',
    'equilibrio',
```



'el banco de mexico, con fundamento en los articulos 8o. y 10o. del  
 ↳reglamento interior del banco de mexico y de conformidad con el  
 ↳procedimiento establecido en el capitulo iv del titulo tercero de su  
 ↳circular 3/2012, informa que la tasa de interes interbancaria de equilibrio  
 ↳en moneda nacional (tiie) a plazo de 28 dias obtenida el dia de hoy, fue de  
 ↳11.2400%; a plazo de 91 dias obtenida el dia de hoy, fue de 11.4029%; y a  
 ↳plazo de 182 dias obtenida el dia de hoy, fue de 11.5604%. la tasa de  
 ↳interes a plazo de 28 dias se calculo con base en las cotizaciones  
 ↳presentadas por las siguientes instituciones de banca multiple: bbva mexico,  
 ↳s.a., banco santander (mexico), s.a., hsbc mexico, s.a., banco nacional de  
 ↳mexico, s.a., banco inbursa, s.a., banca mifel, s.a. y banco j.p. morgan, s.  
 ↳a. ciudad de mexico, a 7 de mayo de 2024.- banco de mexico: directora de  
 ↳disposiciones de banca central, lic. maria teresa munoz aramburu.- rubrica.-  
 ↳gerente de analisis de mercados nacionales, lic. dafne ramos ruiz.- rubrica.  
 ↳tasa de interes interbancaria de equilibrio de fondeo a un dia habil  
 ↳bancario. al margen un logotipo, que dice: banco de mexico.- "2024, ano de  
 ↳felipe carrillo puerto, benemerito del proletariado, revolucionario y  
 ↳defensor del mayab". tasa de interes interbancaria de equilibrio de fondeo a  
 ↳un dia habil bancario',

'el banco de mexico, con fundamento en los articulos 8o. y 10o. del  
 ↳reglamento interior del banco de mexico y de conformidad con el  
 ↳procedimiento establecido en el capitulo iv del titulo tercero de su  
 ↳circular 3/2012, informa que la tasa de interes interbancaria de equilibrio  
 ↳(tiie) de fondeo a un dia habil bancario en moneda nacional determinada el  
 ↳dia de hoy, fue de 11.01 por ciento. ciudad de mexico, a 6 de mayo de 2024.-  
 ↳banco de mexico: directora de disposiciones de banca central, lic. maria  
 ↳teresa munoz aramburu.- rubrica.- gerente de analisis de mercados  
 ↳nacionales, lic. dafne ramos ruiz.- rubrica.'

]

```
def extraer_palabras_clave_tfidf(textos):
    vectorizer = TfidfVectorizer(stop_words=stopwords.words('spanish'))
    tfidf_matrix = vectorizer.fit_transform(textos)
    palabras_clave = []
    for i in range(tfidf_matrix.shape[0]):
        tfidf_scores = zip(vectorizer.get_feature_names_out(), tfidf_matrix[i].
↳toarray()[0])
        sorted_tfidf_scores = sorted(tfidf_scores, key=lambda x: x[1],
↳reverse=True)
        palabras_clave.append([word for word, score in sorted_tfidf_scores[:5]])
    return palabras_clave

# Obtener las palabras clave de cada documento
palabras_clave_documentos = extraer_palabras_clave_tfidf(textos)

# Imprimir las palabras clave de cada documento
for i, palabras_clave in enumerate(palabras_clave_documentos, start=1):
```

```

print(f"Documento {i}: {' '.join(palabras_clave)}")

def eliminarSWE(tokens):
    tokensLimpios = []
    stoppalabras_espanol = set(stopwords.words('spanish'))
    for oracion in tokens:
        oracion_limpia = [palabra for palabra in oracion if palabra.lower() not
↪in stoppalabras_espanol]
        tokensLimpios.append(oracion_limpia)
    return tokensLimpios

def palabrasImpFrec(tokens):
    palabras_importantes = FreqDist([word.lower() for oracion in tokens for
↪word in oracion])
    return palabras_importantes

def resumer(texto):
    oraciones = sent_tokenize(texto)
    palabras = [word_tokenize(oracion) for oracion in oraciones]
    stop_words = set(stopwords.words('spanish'))
    palabras_limpias = eliminarSWE(palabras)
    palabras_importantes = palabrasImpFrec(palabras)
    puntuacion_oracion = {i: sum(palabras_importantes.get(word.lower(), 0) for
↪word in oracion if word.lower() not in stop_words) for i, oracion in
↪enumerate(palabras)}
    oraciones_importantes = sorted(puntuacion_oracion.items(), key=lambda x:
↪x[1], reverse=True)[:1]
    texto_resumido = ' '.join(oraciones[index] for index, _ in
↪oraciones_importantes)

    return texto_resumido

# Resumir cada texto
for i, texto in enumerate(textos, start=1):
    resumen = resumer(texto)
    print(f"Resumen del documento {i}: {resumen}")

```

Documento 1: mexico, banco, cambio, tipo, 2024

Documento 2: mexico, banco, dias, plazo, tasa

Documento 3: mexico, banco, dia, lic, rubrica

Resumen del documento 1: al margen un logotipo, que dice: banco de mexico.-

2024, ano de felipe carrillo puerto, benemerito del proletariado, revolucionario y defensor del mayab.

Resumen del documento 2: la tasa de interes a plazo de 28 dias se calculo con base en las cotizaciones presentadas por las siguientes instituciones de banca multiple: bbva mexico, s.a., banco santander (mexico), s.a., hsbc mexico, s.a., banco nacional de mexico, s.a., banco inbursa, s.a., banca mifel, s.a. y banco

j.p. morgan, s.a. ciudad de mexico, a 7 de mayo de 2024.- banco de mexico:  
directora de disposiciones de banca central, lic.  
Resumen del documento 3: del reglamento interior del banco de mexico y de  
conformidad con el procedimiento establecido en el capitulo iv del titulo  
tercero de su circular 3/2012, informa que la tasa de interes interbancaria de  
equilibrio (tiie) de fondeo a un dia habil bancario en moneda nacional  
determinada el dia de hoy, fue de 11.01 por ciento.

```
[nltk_data] Downloading package punkt to  
[nltk_data]      C:\Users\ianca\AppData\Roaming\nltk_data...  
[nltk_data]   Package punkt is already up-to-date!  
[nltk_data] Downloading package stopwords to  
[nltk_data]      C:\Users\ianca\AppData\Roaming\nltk_data...  
[nltk_data]   Package stopwords is already up-to-date!
```

### 1.3 Función `sent_tokenize` en Python

La función `sent_tokenize` es una herramienta de procesamiento de lenguaje natural (NLP) en Python que se utiliza para dividir un bloque de texto en oraciones individuales.

#### 1.3.1 Cómo funciona

1. **Importación de la librería:** Primero, necesitas importar la librería adecuada en tu script de Python. La función `sent_tokenize` se encuentra comúnmente en la librería NLTK (Natural Language Toolkit) o en la librería spaCy.
2. **Llamada a la función:** Después de importar la librería, puedes utilizar la función `sent_tokenize` pasando el texto que deseas dividir como argumento.
3. **División del texto:** La función analiza el texto y utiliza reglas gramaticales y de puntuación para identificar los límites entre las oraciones. Luego, devuelve una lista donde cada elemento es una oración del texto original.

#### 1.3.2 Ejemplo de código

```
“python # Importar la librería NLTK y la función sent_tokenize from nltk.tokenize import  
sent_tokenize
```

## 2 Texto de ejemplo

texto = “El procesamiento de lenguaje natural es una disciplina emocionante. Te permite analizar  
y comprender el lenguaje humano de una manera automatizada. Existen muchas técnicas y her-  
ramientas disponibles para realizar tareas de procesamiento de lenguaje natural.”

## 3 Llamar a la función `sent_tokenize`

```
oraciones = sent_tokenize(texto)
```

## 4 Imprimir las oraciones divididas

```
for oracion in oraciones: print(oracion)
```