

Procesamiento del Lenguaje Natural

Rodrigo S. Cortez Madrigal



T15 - Actividad Sumativa 1

```
In [80]: from nltk.tokenize import sent_tokenize, word_tokenize
from nltk.corpus import stopwords
from collections import defaultdict
from string import punctuation
from heapq import nlargest
import markdown
from IPython.display import display, HTML
import nltk

In [81]: nltk.download('stopwords')
nltk.download('punkt_tab')

[nltk_data] Downloading package stopwords to
[nltk_data] /Users/roicort/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt_tab to
[nltk_data] /Users/roicort/nltk_data...
[nltk_data] Package punkt_tab is already up-to-date!

Out[81]: True

In [82]: class FrequencySummarizer:
def __init__(self, min_cut=0.1, max_cut=0.9):
    """
    Initilize the text summarizer.
    Words that have a frequency term lower than min_cut
    or higer than max_cut will be ignored.
    """
    self._min_cut = min_cut
    self._max_cut = max_cut
    self._stopwords = set(stopwords.words('spanish') + list(punctuation))

def _compute_frequencies(self, word_sent):
    """
    Compute the frequency of each of word.
    Input:
    word_sent, a list of sentences already tokenized.
    Output:
    freq, a dictionary where freq[w] is the frequency of w.
    """
    freq = defaultdict(int)
    for s in word_sent:
        for word in s:
            if word not in self._stopwords:
                freq[word] += 1
    # frequencies normalization and filtering
    m = float(max(freq.values()))
    to_remove = []
    for w in freq.keys():
        freq[w] = freq[w]/m
        if freq[w] >= self._max_cut or freq[w] <= self._min_cut:
            to_remove.append(w)
    for w in to_remove:
        del freq[w]
    return freq

def summarize(self, text, n):
    """
    Return a list of n sentences
    which represent the summary of text.
    """
    sents = sent_tokenize(text)
    assert n <= len(sents)
    word_sent = [word_tokenize(s.lower()) for s in sents]
    self._freq = self._compute_frequencies(word_sent)
    ranking = defaultdict(int)
    for i,sent in enumerate(word_sent):
        for w in sent:
            if w in self._freq:
                ranking[i] += self._freq[w]
    sents_idx = self._rank(ranking, n)
    return [sents[j] for j in sents_idx]

def _rank(self, ranking, n):
    """ return the first n sentences with highest ranking """
    return nlargest(n, ranking, key=ranking.get)

In [83]: import urllib
from bs4 import BeautifulSoup

def get_only_text(url):
    """
    return the title and the text of the article
    at the specified url
    """
    page = urllib.request.urlopen(url).read().decode('utf8')
    soup = BeautifulSoup(page)
    text = ' '.join(map(lambda p: p.text, soup.find_all('p')))
    return soup.title.text, text

In [84]: import urllib.request
from bs4 import BeautifulSoup

feed_xml = urllib.request.urlopen('https://feeds.elpais.com/mrss-s/pages/ep/site/elpais.com/section/mexico/portada').read()
feed = BeautifulSoup(feed_xml.decode('utf8'), 'lxml-xml')

to_summarize = map(lambda p: p.text, feed.find_all('guid'))

In [85]: articles = list(to_summarize)
articles[:5]

Out[85]: ['https://elpais.com/internacional/2025-05-19/trump-y-putin-hablan-durante-dos-horas-para-revitalizar-las-negociaciones-de-paz-sobre-ucrania.html',
'https://elpais.com/us/2025-05-19/estados-unidos-abre-una-nueva-y-contradictoria-etapa-en-su-lucha-contr-a-el-narcotrafico-mexicano.html',
'https://elpais.com/mexico/2025-05-19/los-maestros-de-la-cn-te-aumentan-la-presion-contr-a-el-gobierno-y-toman-las-principales-casetas-de-entrada-a-la-capital.html',
'https://elpais.com/mexico/2025-05-19/sheinbaum-afirma-que-su-gobierno-no-impulsa-por-el-momento-la-reforma-sobre-las-fiscalias.html',
'https://elpais.com/mexico/2025-05-19/al-menos-ocho-muertos-y-mas-de-una-decena-de-heridos-la-huella-del-caos-tras-el-acamoto-2025-el-festival-motoquero-en-acapulco.html']

In [88]: fs = FrequencySummarizer()
output_lines = ['# Resúmenes de El País\n']

for article_url in articles[:5]:
    title, text = get_only_text(article_url)
    output_lines.append('## ' + title.strip())
    summary = fs.summarize(text, 2)
    for s in summary:
        output_lines.append('- ' + s.strip())
    output_lines.append('\n')

# Guardar en archivo Markdown
with open('resúmenes.md', 'w', encoding='utf-8') as f:
    f.write('\n'.join(output_lines))

In [89]: with open('resúmenes.md', 'r', encoding='utf-8') as f:
    markdown_text = f.read()

html = markdown.markdown(markdown_text)
display(HTML(html))
```

Resúmenes de El País

Trump asegura tras hablar con Putin y Zelenski que Rusia y Ucrania van a abrir negociaciones para acabar con la guerra | Internacional | EL PAÍS

- Ucrania y Rusia comenzarán “de inmediato” negociaciones para un alto el fuego y para acabar la guerra, según ha asegurado este lunes el presidente de Estados Unidos, Donald Trump, tras conversar por teléfono durante dos horas con su homólogo ruso, Vladimir Putin, y poco antes con el líder ucranio, Volodímir Zelenski.
- “Rusia y Ucrania están en el camino correcto tras la reanudación de sus conversaciones directas”, ha declarado el líder ruso, que reitera a Trump sus promesas de paz mientras sus bombas caen sobre las ciudades ucranias día tras día y se muestra reacio a aceptar un alto el fuego por el temor a que no pueda hacer capitular a Kiev en el futuro.

Estados Unidos abre una nueva y contradictoria etapa en su lucha contra el narcotráfico mexicano | EL PAÍS US

- Lo hace en la misma semana en la que el Gobierno de Donald Trump permitió la entrada al país a 17 personas vinculadas a Ovidio Guzmán López, uno de los hijos de Joaquín El Chapo Guzmán.
- Tampoco la de Inzunza Noriega, de 62 años, e Inzunza Coronel, de 33 años, quienes son considerados prófugos de la justicia y quienes han sido imputados después de la celebración de un gran jurado.

Los maestros de la CNTE aumentan la presión contra el Gobierno y toman las principales casetas de entrada a la capital | EL PAÍS México

- Primero retiró la polémica reforma del ISSSTE; más tarde propuso recuperar a través de un decreto, las medidas más beneficiosas para los trabajadores, como la congelación de la edad de jubilación en 56 y 58 años para mujeres y hombres, respectivamente, una medida que afectaría a en torno al 24% del sector, según los cálculos de la Coordinadora, aquellos que pudieron mantenerse bajo el anterior sistema cuando se aprobó la reforma de 2007.
- Los maestros piden ahora volver al modelo anterior, pero desde el Gobierno alegan que no hay dinero suficiente para modificar de golpe este sistema, y están intentando compensarlo con una batería de propuestas que no requieren un aumento tan grande del presupuesto, como la ampliación de las vacaciones, que no estaba entre las peticiones del magisterio pero con la que pensaban aplacar los ánimos, todavía sin éxito.

Sheinbaum afirma que su Gobierno no impulsa “por el momento” la reforma sobre las fiscalías | EL PAÍS México

- Este lunes, preguntada sobre el tema en la conferencia Mañanera, Sheinbaum ha afirmado que primero se debe consolidar la reforma al Poder Judicial —con la que los mexicanos elegirán a todos sus jueces por voto— para luego dar paso a un cambio en la operación de las fiscalías.
- Hace una década, cuando se aprobó dar independencia a las fiscalías, México celebró lo que parecía ser la entrada a una nueva era en la procuración de la justicia, luego de años de permanente cuestionamiento al trabajo de la PGR, antecesora de la actual Fiscalía General.

Al menos ocho muertos y más de una decena de heridos: la huella del caos tras el Acamoto 2025, el festival motoquero en Acapulco | EL PAÍS México

- El festival motoquero Acamoto 2025, que se realizó del 15 al 18 de mayo, concluyó con un saldo de al menos ocho fallecidos en accidentes, 42 personas detenidas por cometer excesos y actos violentos en vía pública —entre ellas la muerte de una transeúnte—, 290 infracciones y 115 motocicletas decomisadas por falta de documentación, según ha dado a conocer el secretario de Seguridad Pública de ese municipio, Eduardo Bailleres.
- “Hay canijos que se portan muy mal”, justificó Macedonio, anteponiendo la derrama económica para el municipio.