

✓ Ejercicio en clase: Web Scraping

ICCD753 Recuperación de Información – Prof. Iván Carrera (2024-B, EPN-FIS)

Fecha de Entrega: Viernes 24 de enero de 2025

Nombre:

- Dilan Andrade

Este trabajo realiza web scraping para extraer categorías de recetas de la página web de AllRecipes, guarda la información en un archivo CSV y luego procesa las recetas extrayendo su título, descripción, ingredientes y pasos. Utiliza el modelo Word2Vec para generar vectores promedio de las recetas y permite realizar búsquedas semánticas por similitud, comparando una consulta con las recetas y mostrando los resultados más relevantes.

Web Scraping para Extraer Categorías de Recetas

Instalación de Dependencias

Antes de ejecutar el código, asegúrate de instalar las bibliotecas necesarias:

```
!pip install requests beautifulsoup4
```

Explicación:

- 1. Instalación de Paquetes:** Se instalan las bibliotecas necesarias `requests` y `beautifulsoup4` para hacer las solicitudes HTTP y analizar el contenido HTML.
- 2. Realización de la Solicitud HTTP:** Se realiza una solicitud a la URL de la página de AllRecipes con encabezados para simular un navegador real.
- 3. Análisis del HTML:** Usando `BeautifulSoup`, el HTML obtenido se analiza para extraer la información deseada.
- 4. Extracción de Categorías:** Se buscan los elementos `` con la clase correspondiente que contienen los enlaces a las categorías de recetas. Luego, se extraen el nombre y la URL de cada categoría.
- 5. Guardar Datos en CSV:** Los datos extraídos (categorías y URLs) se escriben en un archivo CSV para almacenarlos de manera estructurada.

```
import requests
from bs4 import BeautifulSoup
import csv

# URL de la página web
url = "https://www.allrecipes.com/recipes-a-z-6735880"

# Encabezados adicionales
headers = {
    "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36",
    "Accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8",
    "Accept-Language": "en-US,en;q=0.5",
    "Referer": "https://www.allrecipes.com/"
}

# Realizar la solicitud a la página
response = requests.get(url, headers=headers)
response.raise_for_status() # Verificar que la solicitud fue exitosa

# Analizar el contenido HTML
soup = BeautifulSoup(response.text, 'html.parser')

# Buscar todos los elementos <li> con la clase específica
list_items = soup.find_all('li', class_="mntl-link-list__item")

# Lista para almacenar los datos
data = []

# Extraer la categoría y el enlace de cada elemento
for item in list_items:
    link_tag = item.find('a', class_="mntl-link-list__link")
    if link_tag and 'href' in link_tag.attrs:
        categoria = link_tag.text.strip() # Nombre de la categoría
        url = link_tag['href'] # URL
        data.append({'categoria': categoria, 'url': url})

# Guardar los datos en un archivo CSV
csv_file = "categorias_recetas.csv"
```

```

with open(csv_file, mode='w', newline='', encoding='utf-8') as file:
    writer = csv.DictWriter(file, fieldnames=['categoria', 'url'])
    writer.writeheader()
    writer.writerows(data)

print(f"Datos guardados en {csv_file} exitosamente.")

```

 Datos guardados en categorias_recetas.csv exitosamente.

✓ Extracción de Recetas desde Categorías Específicas

Leemos un archivo CSV con categorías y URLs previamente extraídas, luego realiza scraping sobre cada una de esas URLs para obtener información detallada sobre las recetas, como su ID, nombre y URL. Los datos son guardados en un archivo CSV sin duplicados.

Explicación:

1. **Lectura del Archivo CSV de Categorías:** Se lee el archivo CSV que contiene las categorías y URLs extraídas previamente.
2. **Solicitud HTTP a Cada URL de Categoría:** Para cada URL de categoría, se realiza una solicitud HTTP con los mismos encabezados para obtener el contenido de la página de recetas.
3. **Análisis de las Páginas de Recetas:** Usando `BeautifulSoup`, se extraen los enlaces de las recetas dentro de la página de cada categoría.
4. **Extracción de Información de Recetas:** Se obtienen el ID de la receta, su nombre y la URL de cada receta encontrada.
5. **Evitar Duplicados:** Se asegura que cada receta sea única verificando su ID antes de agregarla a la lista de datos.
6. **Guardar los Datos en CSV:** Finalmente, los datos de las recetas (categoría, ID, nombre y URL) se guardan en un archivo CSV sin duplicados.

```

import requests
from bs4 import BeautifulSoup
import csv

# Archivo CSV de entrada (categorías y URLs)
input_csv = "categorias_recetas.csv"

# Archivo CSV de salida (datos de las recetas)
output_csv = "recetas_extraidas.csv"

# Encabezados para el archivo de salida
output_headers = ['categoria', 'id_receta', 'nombre_receta', 'url_receta']

# Encabezados HTTP para la solicitud
headers = {
    "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36",
    "Accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8",
    "Accept-Language": "en-US,en;q=0.5",
}

# Set para almacenar IDs únicos de recetas y evitar duplicados
unique_ids = set()

# Lista para almacenar los datos extraídos
recipes_data = []

# Leer el archivo de categorías y procesar cada URL
with open(input_csv, mode='r', encoding='utf-8') as file:
    reader = csv.DictReader(file)

    for row in reader:
        categoria = row['categoria']
        url = row['url']

        print(f"Procesando categoría: {categoria} - URL: {url}")

        # Realizar solicitud a la URL de la categoría
        response = requests.get(url, headers=headers)
        response.raise_for_status() # Verificar que la solicitud fue exitosa

        # Analizar el contenido HTML de la página
        soup = BeautifulSoup(response.text, 'html.parser')

        # Buscar todas las recetas en la página
        recipes = soup.find_all('a', class_='mntl-card-list-items')

        for recipe in recipes:
            recipe_url = recipe['href'] # URL de la receta

```

```

recipe_id = recipe.get('data-doc-id') # ID de la receta
recipe_name = recipe.find('span', class_='card_title-text').text.strip() # Nombre de la receta

# Evitar duplicados verificando el ID de la receta
if recipe_id not in unique_ids:
    unique_ids.add(recipe_id) # Agregar el ID al set de únicos
    recipes_data.append({
        'categoria': categoria,
        'id_receta': recipe_id,
        'nombre_receta': recipe_name,
        'url_receta': recipe_url
    })

# Guardar los datos extraídos en un nuevo archivo CSV
with open(output_csv, mode='w', newline='', encoding='utf-8') as file:
    writer = csv.DictWriter(file, fieldnames=output_headers)
    writer.writeheader()
    writer.writerows(recipes_data)

print(f"Datos de recetas guardados en {output_csv} exitosamente.")

```

Procesando categoría: Pork Tenderloin - URL: <https://www.allrecipes.com/recipes/1014/meat-and-poultry/pork/tenderloin/>

Procesando categoría: Pot Pies - URL: <https://www.allrecipes.com/recipes/2268/main-dish/savory-pies/pot-pie/>

Procesando categoría: Pot Roast - URL: <https://www.allrecipes.com/recipes/1469/main-dish/beef/pot-roast/>

Procesando categoría: Potato Pancakes - URL: <https://www.allrecipes.com/recipes/1545/side-dish/potatoes/potato-pancakes/>

Procesando categoría: Potato Salads - URL: <https://www.allrecipes.com/recipes/217/salad/potato-salad/>

Procesando categoría: Potato Soups - URL: <https://www.allrecipes.com/recipes/1552/soups-stews-and-chili/soup/vegetable-soup/potat>

Procesando categoría: Potluck Recipes - URL: <https://www.allrecipes.com/recipes/1536/everyday-cooking/on-the-go/potluck/>

Procesando categoría: Pound Cakes - URL: <https://www.allrecipes.com/recipes/395/desserts/cakes/pound-cake/>

Procesando categoría: Pretzels - URL: <https://www.allrecipes.com/recipes/16220/bread/yeast-bread/pretzels/>

Procesando categoría: Prime Rib - URL: <https://www.allrecipes.com/recipes/1012/meat-and-poultry/beef/rib-roast/>

Procesando categoría: Pulled Pork - URL: <https://www.allrecipes.com/recipes/16543/main-dish/pork/pulled-pork/>

Procesando categoría: Pumpkin Breads - URL: <https://www.allrecipes.com/recipes/347/bread/quick-bread/pumpkin-bread/>

Procesando categoría: Pumpkin Pie - URL: <https://www.allrecipes.com/recipes/819/desserts/pies/pumpkin-pie/>

Procesando categoría: Pumpkin Seeds - URL: <https://www.allrecipes.com/recipes/969/appetizers-and-snacks/nuts-and-seeds/pumpkin-se>

Procesando categoría: Punch - URL: <https://www.allrecipes.com/recipes/136/drinks/punch/>

Procesando categoría: Purim - URL: <https://www.allrecipes.com/recipes/1444/holidays-and-events/purim/>

Procesando categoría: Quesadillas - URL: <https://www.allrecipes.com/recipes/1905/world-cuisine/latin-american/mexican/main-dishes>

Procesando categoría: Quiches - URL: <https://www.allrecipes.com/recipes/16793/main-dish/quiche/>

Procesando categoría: Quick and Easy Recipes - URL: <https://www.allrecipes.com/recipes/1947/everyday-cooking/quick-and-easy/>

Procesando categoría: Quinoa - URL: <https://www.allrecipes.com/recipes/15592/ingredients/whole-grains/quinoa/>

Procesando categoría: Ramadan - URL: <https://www.allrecipes.com/recipes/195/holidays-and-events/ramadan/>

Procesando categoría: Ravioli - URL: <https://www.allrecipes.com/recipes/16855/pasta-and-noodles/pasta-by-shape/ravioli/>

Procesando categoría: Raw Food Diet - URL: <https://www.allrecipes.com/recipes/16337/healthy-recipes/raw-food-diet/>

Procesando categoría: Refried Beans - URL: <https://www.allrecipes.com/recipes/17873/side-dish/beans-and-peas/refried-beans/>

Procesando categoría: Relishes - URL: <https://www.allrecipes.com/recipes/13378/side-dish/sauces-and-condiments/relishes/>

Procesando categoría: Rhubarb Pie - URL: <https://www.allrecipes.com/recipes/12953/desserts/pies/fruit-pies/rhubarb-pie/>

Procesando categoría: Ribs - URL: <https://www.allrecipes.com/recipes/261/main-dish/ribs/>

Procesando categoría: Rice Casseroles - URL: <https://www.allrecipes.com/recipes/1004/main-dish/casseroles/rice/>

Procesando categoría: Rice Pilaf - URL: <https://www.allrecipes.com/recipes/14364/side-dish/rice/pilaf/>

Procesando categoría: Rice Puddings - URL: <https://www.allrecipes.com/recipes/14442/desserts/custards-and-puddings/rice-pudding/>

Procesando categoría: Risotto - URL: <https://www.allrecipes.com/recipes/14362/main-dish/rice/risotto/>

Procesando categoría: Road Trip Snacks - URL: <https://www.allrecipes.com/recipes/1991/everyday-cooking/on-the-go/road-trip-snacks>

Procesando categoría: Roasts - URL: <https://www.allrecipes.com/recipes/258/main-dish/roasts/>

Procesando categoría: Rosh Hashanah - URL: <https://www.allrecipes.com/recipes/196/holidays-and-events/rosh-hashanah/>

Procesando categoría: Salad Dressings - URL: <https://www.allrecipes.com/recipes/210/side-dish/sauces-and-condiments/salad-dressin>

Procesando categoría: Salads - URL: <https://www.allrecipes.com/recipes/96/salad/>

Procesando categoría: Salisbury Steak - URL: <https://www.allrecipes.com/recipes/17656/main-dish/beef/salisbury-steak/>

Procesando categoría: Salmon - URL: <https://www.allrecipes.com/recipes/416/seafood/fish/salmon/>

Procesando categoría: Salsas - URL: <https://www.allrecipes.com/recipes/436/appetizers-and-snacks/dips-and-spreads/salsa/>

Procesando categoría: Samosas - URL: <https://www.allrecipes.com/recipes/17776/appetizers-and-snacks/pastries/samosas/>

Procesando categoría: Sandwich Cookies - URL: <https://www.allrecipes.com/recipes/16416/desserts/cookies/sandwich-cookies/>

Procesando categoría: Sandwiches - URL: <https://www.allrecipes.com/recipes/251/main-dish/sandwiches/>

Procesando categoría: Sangrias - URL: <https://www.allrecipes.com/recipes/242/drinks/sangria/>

Procesando categoría: Sausage - URL: <https://www.allrecipes.com/recipes/16546/meat-and-poultry/sausage/>

Procesando categoría: Scones - URL: <https://www.allrecipes.com/recipes/1997/bread/quick-bread/scones/>

Procesando categoría: Seder - URL: <https://www.allrecipes.com/recipes/985/holidays-and-events/passover/seder/>

Procesando categoría: Seitan - URL: <https://www.allrecipes.com/recipes/17624/everyday-cooking/vegetarian/protein/seitan/>

Procesando categoría: Shepherd's Pie - URL: <https://www.allrecipes.com/recipes/2270/main-dish/savory-pies/shepherds-pie/>

Procesando categoría: Shortbread - URL: <https://www.allrecipes.com/recipes/855/desserts/cookies/butter-cookies/shortbread-cookies>

Procesando categoría: Shortcakes - URL: <https://www.allrecipes.com/recipes/15066/desserts/cakes/shortcake/>

Procesando categoría: Shrimp and Grits - URL: <https://www.allrecipes.com/recipes/16685/main-dish/seafood/shrimp/shrimp-and-grits/>

Procesando categoría: Shrimp Scampi - URL: <https://www.allrecipes.com/recipes/15853/main-dish/seafood/shrimp/shrimp-scampi/>

Procesando categoría: Side Dishes - URL: <https://www.allrecipes.com/recipes/81/side-dish/>

Procesando categoría: Slab Pie - URL: <https://www.allrecipes.com/recipes/17640/desserts/pies/slab-pie/>

Procesando categoría: Sliders - URL: <https://www.allrecipes.com/recipes/17308/everyday-cooking/more-ideas/mini-foods/main-di>

Procesando categoría: Sloppy Joes - URL: <https://www.allrecipes.com/recipes/14930/main-dish/sandwiches/sloppy-joes/>

Procesando categoría: Slow Cooker Recipes - URL: <https://www.allrecipes.com/recipes/253/everyday-cooking/slow-cooker/>

✓ Extracción Detallada de Recetas desde Enlaces

Este script extrae información detallada de recetas (título, descripción, ingredientes y pasos) desde URLs almacenadas en un archivo CSV y guarda los datos en un nuevo archivo CSV, procesando solo los primeros 1000 enlaces.

Explicación:

1. **Lectura del CSV de Recetas:** Se lee un archivo CSV con URLs de recetas.
2. **Solicitud HTTP y Análisis HTML:** Para cada URL, se realiza una solicitud HTTP y se analiza el contenido HTML para extraer título, descripción, ingredientes y pasos.
3. **Validación de Datos:** Se verifica que todos los campos estén completos antes de guardar los datos.
4. **Guardar en CSV:** Los datos extraídos se almacenan en un archivo CSV, con ingredientes y pasos en formato de lista.

```
import requests
from bs4 import BeautifulSoup
import csv

# Archivo de entrada con las URLs de las recetas
input_csv = "recetas_extraidas.csv"

# Archivo de salida con los datos procesados
output_csv = "Recetas_DilanAndrade.csv"

# Encabezados del archivo de salida
output_headers = ['categoria', 'titulo_receta', 'descripcion_receta', 'ingredientes', 'pasos']

# Encabezados HTTP para la solicitud
headers = {
    "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36",
    "Accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8",
    "Accept-Language": "en-US,en;q=0.5",
}

# Lista para almacenar los datos extraídos
recipes_details = []

# Leer el archivo CSV de recetas y procesar los primeros 1000 enlaces
with open(input_csv, mode='r', encoding='utf-8') as file:
    reader = csv.DictReader(file)
    for i, row in enumerate(reader):
        if i >= 1000: # Limitar a los primeros 1000 enlaces
            break

        categoria = row['categoria']
        url_receta = row['url_receta']

        print(f"Procesando receta {i + 1}: {url_receta}")

        try:
            # Realizar la solicitud a la página de la receta
            response = requests.get(url_receta, headers=headers)
            response.raise_for_status() # Verificar que la solicitud fue exitosa

            # Analizar el contenido HTML
            soup = BeautifulSoup(response.text, 'html.parser')

            # Extraer título de la receta
            titulo_receta_tag = soup.find('h1', class_='article-heading text-headline-400')
            titulo_receta = titulo_receta_tag.text.strip() if titulo_receta_tag else None

            # Extraer descripción de la receta
            descripcion_receta_tag = soup.find('p', class_='article-subheading text-body-100')
            descripcion_receta = descripcion_receta_tag.text.strip() if descripcion_receta_tag else None

            # Extraer ingredientes como una lista (línea por línea)
            ingredientes_list = soup.find_all('li', class_='mm-recipes-structured-ingredients__list-item')
            ingredientes = [item.text.strip() for item in ingredientes_list] if ingredientes_list else None

            # Extraer pasos de la receta
            pasos_list = soup.find_all('li', class_='comp mntl-sc-block mntl-sc-block-startgroup mntl-sc-block-group--LI')
            pasos = [step.find('p', class_='comp mntl-sc-block mntl-sc-block-html').text.strip() for step in pasos_list] if pasos_list else None

            # Validar que todos los datos requeridos existan
            if titulo_receta and descripcion_receta and ingredientes and pasos:
                # Guardar los datos extraídos
                recipes_details.append({
                    'categoria': categoria,
                    'titulo_receta': titulo_receta,
                    'descripcion_receta': descripcion_receta,
                    'ingredientes': ingredientes, # Guardar como lista directamente
                    'pasos': pasos # Guardar los pasos como lista
                })
            else:
                print(f"Receta {i + 1} no tiene todos los datos requeridos.")
```

```

print('Datos incompletos en la receta {i + 1}, no se agregará. ')

except Exception as e:
    print(f'Error procesando la receta {i + 1}: {e}')
    continue

# Guardar los detalles en un archivo CSV
with open(output_csv, mode='w', newline='', encoding='utf-8') as file:
    writer = csv.writer(file)
    writer.writerow(output_headers)

    for recipe in recipes_details:
        # Escribir la receta en el archivo
        writer.writerow([
            recipe['categoria'],
            recipe['titulo_receta'],
            recipe['descripcion_receta'],
            "\n".join(recipe['ingredientes']), # Cada ingrediente en una nueva línea
            "\n".join(recipe['pasos']) # Cada paso en una nueva línea
        ])

print(f"Datos de las recetas guardados en {output_csv} exitosamente.")

```

Procesando receta 1: <https://www.allrecipes.com/gallery/best-air-fryer-snack-recipes/>
 Datos incompletos en la receta 1, no se agregará.
 Procesando receta 2: <https://www.allrecipes.com/gallery/air-fryer-side-dishes-for-dinner/>
 Datos incompletos en la receta 2, no se agregará.
 Procesando receta 3: <https://www.allrecipes.com/article/how-to-convert-recipes-for-an-air-fryer/>
 Datos incompletos en la receta 3, no se agregará.
 Procesando receta 4: <https://www.allrecipes.com/article/air-fryer-tips/>
 Datos incompletos en la receta 4, no se agregará.
 Procesando receta 5: <https://www.allrecipes.com/ninja-crispi-air-fryer-amazon-2024-8732423>
 Datos incompletos en la receta 5, no se agregará.
 Procesando receta 6: <https://www.allrecipes.com/recipe/283850/easy-air-fried-chicken-breast/>
 Procesando receta 7: <https://www.allrecipes.com/air-fryer-appetizer-recipes-8746703>
 Datos incompletos en la receta 7, no se agregará.
 Procesando receta 8: <https://www.allrecipes.com/air-fryer-lemon-garlic-parmesan-chicken-recipe-8726749>
 Procesando receta 9: <https://www.allrecipes.com/best-air-fryer-thanksgiving-recipes-8729089>
 Datos incompletos en la receta 9, no se agregará.
 Procesando receta 10: <https://www.allrecipes.com/air-fryer-s-mores-recipe-8736955>
 Procesando receta 11: <https://www.allrecipes.com/air-fryer-baked-yams-recipe-8737640>
 Procesando receta 12: <https://www.allrecipes.com/lemon-garlic-butter-chicken-spiedini-recipe-8727930>
 Procesando receta 13: <https://www.allrecipes.com/air-fryer-grilled-pimento-cheese-recipe-8720986>
 Procesando receta 14: <https://www.allrecipes.com/air-fryer-chicken-parmesan-recipe-8698442>
 Procesando receta 15: <https://www.allrecipes.com/air-fryer-eggplant-recipe-8700773>
 Procesando receta 16: <https://www.allrecipes.com/air-fryer-sriracha-fries-recipe-8694021>
 Procesando receta 17: <https://www.allrecipes.com/crispy-air-fryer-potato-bites-recipe-8682451>
 Procesando receta 18: <https://www.allrecipes.com/copycat-wingstop-cajun-corn-recipe-8681405>
 Procesando receta 19: <https://www.allrecipes.com/air-fryer-po-boy-recipe-8647044>
 Procesando receta 20: <https://www.allrecipes.com/air-fryer-buffalo-wings-recipe-8645970>
 Procesando receta 21: <https://www.allrecipes.com/air-fryer-smashed-potatoes-recipe-8644582>
 Procesando receta 22: <https://www.allrecipes.com/air-fryer-quesadillas-recipe-8651442>
 Procesando receta 23: <https://www.allrecipes.com/air-fryer-truffle-polenta-fries-recipe-8643717>
 Procesando receta 24: <https://www.allrecipes.com/air-fryer-firecracker-salmon-bites-recipe-8623119>
 Procesando receta 25: <https://www.allrecipes.com/air-fryer-chicken-bites-recipe-8599352>
 Procesando receta 26: <https://www.allrecipes.com/4-ingredient-air-fryer-pepper-poppers-recipe-8584965>
 Procesando receta 27: <https://www.allrecipes.com/air-fryer-bell-pepper-poppers-recipe-8584648>
 Procesando receta 28: <https://www.allrecipes.com/air-fryer-cinnamon-roll-bites-recipe-8553617>
 Procesando receta 29: <https://www.allrecipes.com/air-fryer-ham-and-cheese-wraps-recipe-8365118>
 Procesando receta 30: <https://www.allrecipes.com/recipe/8537905/air-fryer-buffalo-cauliflower/>
 Procesando receta 31: <https://www.allrecipes.com/air-fryer-honey-mustard-chicken-thighs-recipe-7970816>
 Procesando receta 32: <https://www.allrecipes.com/air-fryer-hearts-of-palm-sticks-recipe-7814346>
 Procesando receta 33: <https://www.allrecipes.com/air-fryer-cheesy-bacon-ranch-french-fries-recipe-7560981>
 Procesando receta 34: <https://www.allrecipes.com/air-fryer-mini-dark-chocolate-cake-with-brown-butter-frosting-recipe-7509612>
 Procesando receta 35: <https://www.allrecipes.com/air-fryer-spanakopita-recipe-7510155>
 Procesando receta 36: <https://www.allrecipes.com/air-fryer-pecan-crusted-trout-recipe-7510184>
 Procesando receta 37: <https://www.allrecipes.com/air-fryer-blooming-onion-recipe-7546676>
 Procesando receta 38: <https://www.allrecipes.com/air-fryer-tempura-vegetables-recipe-7510181>
 Procesando receta 39: <https://www.allrecipes.com/air-fryer-turtle-cheesecake-recipe-7511366>
 Procesando receta 40: <https://www.allrecipes.com/air-fryer-samosas-recipe-7510128>
 Procesando receta 41: <https://www.allrecipes.com/air-fryer-roasted-vegetables-with-gremolata-and-red-pepper-aioli-recipe-7510058>
 Procesando receta 42: <https://www.allrecipes.com/crab-rangoon-in-the-air-fryer-recipe-7511370>
 Procesando receta 43: <https://www.allrecipes.com/air-fryer-everything-bagel-salmon-dinner-recipe-7509518>
 Procesando receta 44: <https://www.allrecipes.com/air-fryer-crispy-pickle-chips-with-creole-dipping-sauce-recipe-7509784>
 Procesando receta 45: <https://www.allrecipes.com/air-fryer-raspberry-brownie-bites-recipe-7509875>
 Procesando receta 46: <https://www.allrecipes.com/air-fryer-sweet-potato-chips-recipe-7510177>
 Procesando receta 47: <https://www.allrecipes.com/air-fryer-green-bean-fries-recipe-7509535>
 Procesando receta 48: <https://www.allrecipes.com/air-fryer-loaded-egg-salad-sandwiches-recipe-7509591>
 Procesando receta 49: <https://www.allrecipes.com/air-fryer-chicken-tenders-recipe-7509380>
 Procesando receta 50: <https://www.allrecipes.com/air-fryer-garlic-green-beans-recipe-7511002>

✓ Preparación del corpus

Importamos las bibliotecas necesarias y descargamos los recursos de NLTK para tokenización y eliminación de stopwords, además de cargar un modelo de palabras preentrenado.

Explicación:

1. **Importamos las Bibliotecas:** Cargamos pandas, numpy, nltk, y gensim para procesamiento de texto y modelos preentrenados.
2. **Descargamos los Recursos NLTK:** Bajamos los recursos para tokenización y stopwords.

```
import pandas as pd
import numpy as np
import re
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import nltk
import gensim.downloader as api
from numpy.linalg import norm
from tabulate import tabulate # Para mostrar los resultados en una tabla
```

```
# Descargar recursos de NLTK
```

```
nltk.download('punkt')
nltk.download('punkt_tab')
nltk.download('stopwords')
```

```

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt_tab.zip.
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
True
```

✓ Funciones para Procesamiento de Texto y Búsqueda de Similitudes

Implementamos funciones para preprocesar texto, generar vectores promedio, calcular similitud entre vectores usando el coseno, y buscar recetas relevantes basadas en una consulta.

Explicación:

1. **Preprocesamiento de Texto:**
 - Eliminamos caracteres especiales y convertimos el texto a minúsculas.
 - Tokenizamos el texto en palabras.
2. **Generación de Vectores Promedio:**
 - Obtenemos el vector promedio de las palabras presentes en el modelo de palabras preentrenadas.
 - Si no se encuentran palabras válidas, devolvemos un vector nulo.
3. **Similitud del Coseno:**
 - Calculamos la similitud entre dos vectores usando el coseno, retornando un valor entre 0 y 1.
4. **Búsqueda de Recetas Relevantes:**
 - Generamos el vector de la consulta, calculamos la similitud con los vectores de las recetas y ordenamos los resultados por similitud para devolver las mejores coincidencias.

```
# Función de preprocesamiento
def preprocess(text):
    text = re.sub(r'[^w\s]', '', text.lower()) # Eliminar caracteres especiales y convertir a minúsculas
    tokens = word_tokenize(text) # Tokenizar

    return tokens
```

```
# Función para generar el vector promedio de un texto
def get_vector(text, model):
    tokens = preprocess(text)
    in_vocab = [word for word in tokens if word in model]
    print(f"Tokens en vocabulario: {len(in_vocab)} de {len(tokens)}") # Debug
    vectors = [model[word] for word in in_vocab]
    if vectors:
        return np.mean(vectors, axis=0)
    else:
        return np.zeros(model.vector_size) # Vector nulo si no hay palabras válidas
```

```
# Función de similitud del coseno
def cosine_similarity(vec1, vec2):
    norm1 = norm(vec1)
```

```

norm2 = norm(vec2)
if norm1 == 0 or norm2 == 0:
    return 0.0
return np.dot(vec1, vec2) / (norm1 * norm2)

# Función para buscar recetas relevantes
def search(query, data, model, top_n=5):
    query_vector = get_vector(query, model) # Generar vector de la consulta
    data['similarity'] = data['vector'].apply(lambda vec: cosine_similarity(query_vector, vec))
    results = data.sort_values(by='similarity', ascending=False).head(top_n) # Ordenar por similitud
    return results[['categoria', 'titulo_receta', 'descripcion_receta', 'similarity']]

```

✓ Carga del Modelo Word2Vec

Cargamos el modelo preentrenado de Word2Vec ("word2vec-google-news-300") para utilizarlo en tareas de análisis semántico, como generar vectores de palabras y calcular similitudes.

```

print("Cargando el modelo Word2Vec...")
word2vec_model = api.load("word2vec-google-news-300")
print("Modelo cargado exitosamente.")

```

```

↗ Cargando el modelo Word2Vec...
[=====] 100.0% 1662.8/1662.8MB downloaded
Modelo cargado exitosamente.

```

✓ Carga y Preprocesamiento de Datos de Recetas

Cargamos un archivo CSV con datos de recetas y creamos una nueva columna que combina el título, descripción, ingredientes y pasos de cada receta para formar un texto completo.

Explicación por Pasos:

1. Cargar el CSV:

- Usamos `pandas` para leer el archivo CSV que contiene los datos de las recetas.

2. Crear Columna de Texto Completo:

- Combinamos el título, descripción, ingredientes y pasos de cada receta en una nueva columna llamada `texto_completo`.

3. Verificar Contenido:

- Mostramos las primeras filas para verificar que los datos se han combinado correctamente.

```

file_path = "/content/Recetas_DilanAndrade.csv" # Cambia por la ruta de tu archivo
print("Cargando el archivo CSV...")
data = pd.read_csv(file_path)
print("Archivo cargado exitosamente.")

# Crear una nueva columna 'texto_completo' combinando las partes relevantes
data['texto_completo'] = data.apply(lambda row: f"{row['titulo_receta']} {row['descripcion_receta']} {row['ingredientes']} {row['pasos']}",
                                   axis=1)

# Verificar el contenido
print(data[['categoria', 'titulo_receta', 'texto_completo']].head())

```

```

↗ Cargando el archivo CSV...
Archivo cargado exitosamente.

```

	categoria	titulo_receta \
0	Air Fryer Recipes	Easy Air Fryer Chicken Breast
1	Air Fryer Recipes	Air Fryer Lemon Garlic Parmesan Chicken
2	Air Fryer Recipes	Air Fryer S'Mores
3	Air Fryer Recipes	Air Fryer Baked Yams
4	Air Fryer Recipes	Lemon Garlic Butter Chicken Spiedini

	texto_completo
0	Easy Air Fryer Chicken Breast This air fryer c...
1	Air Fryer Lemon Garlic Parmesan Chicken These ...
2	Air Fryer S'Mores This recipe for air fryer s'...
3	Air Fryer Baked Yams These air fryer baked yam...
4	Lemon Garlic Butter Chicken Spiedini These lem...

✓ Generación de Vectores para las Recetas

Generamos vectores promedio para cada receta utilizando el modelo Word2Vec cargado, basándonos en el texto completo de cada receta.

Explicación:

1. Generar Vectores:

- Aplicamos la función `get_vector` a la columna `texto_completo` para obtener un vector promedio para cada receta.

2. Almacenar Vectores:

- Los vectores generados se almacenan en una nueva columna llamada `vector`.

```
print("Generando vectores para las recetas...")
data['vector'] = data['texto_completo'].apply(lambda text: get_vector(text, word2vec_model))
print("Vectores generados exitosamente.")
```

```

Tokens en vocabulario: 171 de 189
Tokens en vocabulario: 225 de 267
Tokens en vocabulario: 156 de 181
Tokens en vocabulario: 167 de 185
Tokens en vocabulario: 260 de 291
Tokens en vocabulario: 125 de 147
Tokens en vocabulario: 395 de 445
Tokens en vocabulario: 232 de 258
Tokens en vocabulario: 105 de 119
Tokens en vocabulario: 365 de 416
Tokens en vocabulario: 178 de 199
Tokens en vocabulario: 194 de 218
Tokens en vocabulario: 218 de 242
Tokens en vocabulario: 399 de 473
Tokens en vocabulario: 197 de 221
Tokens en vocabulario: 183 de 212
Tokens en vocabulario: 211 de 243
Tokens en vocabulario: 219 de 255
Tokens en vocabulario: 355 de 409
Tokens en vocabulario: 171 de 190
Tokens en vocabulario: 201 de 226
Tokens en vocabulario: 328 de 371
Tokens en vocabulario: 259 de 297
Tokens en vocabulario: 96 de 108
Tokens en vocabulario: 150 de 176
Tokens en vocabulario: 211 de 257
Tokens en vocabulario: 222 de 249
Tokens en vocabulario: 114 de 134
Tokens en vocabulario: 273 de 322
Tokens en vocabulario: 232 de 272
Tokens en vocabulario: 368 de 437
Tokens en vocabulario: 219 de 251
Tokens en vocabulario: 219 de 247
Tokens en vocabulario: 306 de 351
Tokens en vocabulario: 196 de 224
Tokens en vocabulario: 127 de 145
Tokens en vocabulario: 196 de 229
Tokens en vocabulario: 196 de 224
Tokens en vocabulario: 213 de 238
Tokens en vocabulario: 158 de 191
Tokens en vocabulario: 232 de 269
Tokens en vocabulario: 113 de 121
Tokens en vocabulario: 312 de 369
Tokens en vocabulario: 310 de 355
Tokens en vocabulario: 246 de 284
Tokens en vocabulario: 153 de 177
Tokens en vocabulario: 180 de 212
Tokens en vocabulario: 194 de 224
Tokens en vocabulario: 187 de 215
Tokens en vocabulario: 176 de 195
Tokens en vocabulario: 141 de 166
Tokens en vocabulario: 238 de 264
Tokens en vocabulario: 195 de 216
Tokens en vocabulario: 264 de 297
Tokens en vocabulario: 135 de 154
Tokens en vocabulario: 146 de 167
Tokens en vocabulario: 198 de 226
Tokens en vocabulario: 129 de 144
Tokens en vocabulario: 157 de 181

```

✓ Búsqueda de Recetas por Consulta

Buscamos recetas relacionadas con una consulta (en este caso, "chicken") y mostramos los resultados ordenados por similitud con el término de búsqueda.

Explicación:

1. Realizar Búsqueda:

- Usamos la función `search` para obtener las recetas más relevantes para la consulta dada, ordenadas por similitud.

2. Mostrar Resultados:

- Convertimos los resultados en una tabla y los mostramos utilizando `tabulate` con formato bonito.
- También mostramos la similitud promedio entre las recetas y la consulta.

```
query = "chicken" # Cambia por tu consulta
print(f"Buscando recetas para la consulta: '{query}'...")

# Buscar todas las recetas ordenadas por similitud
results = search(query, data, word2vec_model, top_n=len(data)) # Usar todas las filas disponibles

# Mostrar los resultados en formato tabular
print("\nResultados detallados (todas las recetas):")
headers = ["Categoría", "Título de la Receta", "Descripción", "Similitud"]
table = results.values.tolist() # Convertir los resultados a una lista de listas
print(tabulate(table, headers=headers, tablefmt="fancy_grid", numalign="center")) # Tabular con formato

# Mostrar la similitud promedio
print("\nSimilitud promedio:", round(results['similarity'].mean(), 4))
```

Applesauce	Rhubarb Applesauce	Rhubarb applesauce is tradition
Bagels	Moist Passover Bagel	Moist Passover bagels that are
Biscuits	Spelt Biscuits	I have made these biscuits for
Apple Pie	Easy German Apple Streusel Pie	This apple streusel pie has a s
Air Fryer Recipes	Air Fryer Turtle Cheesecake	Did you know that you can use y
Bar Cookies	Maple Pecan Blondies	These maple pecan blondies are
Banana Breads	Oatmeal Banana Nut Bread	This is the best banana nut-oat
Apple Pie	Mum's Irish Apple Pie	This Irish apple pie recipe is
Biscuits	Canadian Tea Biscuits	These Canadian tea biscuits are
Biscuits	Super Easy Shortcake Biscuits	These shortcake biscuits have b
Bar Cookies	Magic Cookie Bars from Eagle Brand	These magic cookie bars are an
Apple Pie	Upside-Down Apple Pecan Pie	This is by far the best apple p
Applesauce	Applesauce Muffins	These applesauce muffins, with
Apple Pie	Vegan Apple Pie	This vegan apple pie recipe is
Biscuits	Nanny's Newfoundland Tea Biscuits	My nanny used to make these tea
Apple Pie	Chef John's Easy Apple Pie	This apple pie is not your typi
Applesauce	Chunky Cinnamon-Orange Applesauce	A must-have recipe for the Jewi
Bagels	Whole Wheat Bagels	These yummy bagels can be made
Apple Pie	Canned Apple Pie Filling	This apple pie filling recipe i
Apple Pie	Old-Fashioned Apple Pie	This classic and traditional Am
Bar Cookies	Apple Crisp-Browned Butter Bars	Try these delicious apple bars
Biscuits	Homemade Biscuit Mix	Add a few simple ingredients, b
Bagels	Multigrain Bagels	I added flaxseed meal, wheat ge
Apple Pie	Croatian Apple Pie Squares	Croatian apple pie squares feat
Bar Cookies	Chocolate Peanut Butter Protein Bars	Chocolate peanut butter protein
Bar Cookies	Mmm-Mmm Better Brownies	This brownie recipe with oil ma
Apple Pie	Perfect Apple Crisp Pie	Amazing apple crisp pie! My son
Biscotti	Gluten-Free Biscotti	This recipe was given to me by
Biscuits	Sweet Potato Sage Biscuits	These savory sweet potato sage

Start coding or [generate](#) with AI.

