

# Final

December 3, 2023

```
[1]: print("""
Riker Wachtler
3 December 2023
NLP Midterm Project
https://github.com/RikerW/final-nlp
""")
```

Riker Wachtler  
3 December 2023  
NLP Midterm Project  
<https://github.com/RikerW/final-nlp>

```
[22]: print("""
This is the completed version of my final project. I left short comments for_
↪each thing I'm doing - but in short,
it's doing some pre-processing stuff as EDA and then getting into it.

Then it does:
TF-IDF into LDA
NER
WordClouds for a couple reviews for fun
Translation via TextBlob into Spanish
POS tagging
Word2Vec similarity checks
FreqDist from NLTK to find most common words
Sentiment Analysis using pre-trained BertTokenizer

The data is taken from a Rotten Tomatoes review compilation, as a text file.
""")
```

This is the completed version of my final project. I left short comments for each thing I'm doing - but in short, it's doing some pre-processing stuff as EDA and then getting into it.

Then it does:  
TF-IDF into LDA

NER

WordClouds for a couple reviews for fun

Translation via TextBlob into Spanish

POS tagging

Word2Vec similarity checks

FreqDist from NLTK to find most common words

Sentiment Analysis using pre-trained BertTokenizer

The data is taken from a Rotten Tomatoes review compilation, as a text file.

```
[2]: # Just a lot of imports. A lot.

import nltk
from nltk.corpus import stopwords
from nltk.tokenize import sent_tokenize, word_tokenize
from nltk.probability import FreqDist
nltk.download('punkt')
nltk.download('stopwords')

stop_words = set(stopwords.words('english'))

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.decomposition import LatentDirichletAllocation

import re
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline

from wordcloud import WordCloud
from textblob import TextBlob
import gensim

import spacy
from spacy import displacy
nlp = spacy.load('en_core_web_sm')

import random

import warnings
warnings.filterwarnings("ignore")

import tensorflow as tf
from transformers import BertTokenizer, TFBertForSequenceClassification
```

```
C:\Users\riker\anaconda3\lib\site-packages\scipy\__init__.py:155: UserWarning: A
NumPy version >=1.18.5 and <1.25.0 is required for this version of SciPy
(detected version 1.25.0
```

```
warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}")
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\riker\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\riker\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

```
[3]: # basic loading + EDA - the txt file is a tab delimited df with sentences from
      ↪ Rotten Tomatoes reviews for sentiment analysis
df = pd.read_csv("data/datasetSentences.txt", index_col=0, delimiter = "\t")
df = df.dropna()
```

```
[4]: df.head()
```

```
[4]:
```

	sentence
sentence_index	
1	The Rock is destined to be the 21st Century 's...
2	The gorgeously elaborate continuation of `` Th...
3	Effective but too-tepid biopic
4	If you sometimes like to go to the movies to h...
5	Emerges as something rare , an issue movie tha...

```
[5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 11855 entries, 1 to 11855
Data columns (total 1 columns):
#   Column      Non-Null Count  Dtype
---  ---
0    sentence    11855 non-null  object
dtypes: object(1)
memory usage: 185.2+ KB
```

```
[6]: # There's an ungodly amount of reviews, so let's get 500 randomly selected to
      ↪ work with, so we're not dumping massive outputs
random.seed(10)
split = random.randint(0, 10000)
sentence_samples = df[split:split+500]
sentence_samples.reset_index(drop=True, inplace=True)
sentence_samples
```

```
[6]:
```

	sentence
0	Although Barbershop boasts some of today 's ho...
1	A puzzle whose pieces do not fit .

```

2    Some are fascinating and others are not , and ...
3    Would that Greengrass had gone a tad less for ...
4    The good is very , very good ... The rest runs...
..
495 The movie 's plot is almost entirely witless a...
496 ... may work as an addictive guilty pleasure b...
497             This Scarlet 's letter is A. . .
498             as in aimless , arduous , and arbitrary .
499 Plays like a glossy melodrama that occasionall...

```

[500 rows x 1 columns]

```

[7]: # Basic text cleaning function
def preprocess_text(text):
    words = word_tokenize(text)
    cleaned_words = word_tokenize(re.sub(r'[^a-zA-Z]', ' ', text))
    cleaned_words = [word.lower() for word in cleaned_words if word.isalnum()
↳and word not in stop_words]
    return ' '.join(cleaned_words)

```

```

[8]: # Generate a processed version of our sentence samples
sentence_samples_processed = sentence_samples.copy()
for i in range(len(sentence_samples_processed)):
    sentence_samples_processed["sentence"][i] =
↳preprocess_text(sentence_samples_processed["sentence"][i])

sentence_samples_processed

```

```

[8]: sentence
0    although barbershop boasts today hottest hippe...
1             a puzzle whose pieces fit
2    some fascinating others end almost good movie
3    would greengrass gone tad less grit lot intell...
4    the good good the rest runs mildly unimpressiv...
..
495 the movie plot almost entirely witless inane c...
496 may work addictive guilty pleasure material ne...
497             this scarlet letter a
498             aimless arduous arbitrary
499 plays like glossy melodrama occasionally verge...

```

[500 rows x 1 columns]

```

[9]: # TF-IDF vectorize this - and get the scores. used for LDA
tfidf = TfidfVectorizer()
tfidf_matrix = tfidf.fit_transform(sentence_samples_processed["sentence"][0:30])
tfidf_scores = tfidf_matrix.sum(axis=1)

```

```
tfidf_scores
```

```
[9]: matrix([[4.68484218],
            [2.        ],
            [2.63543752],
            [2.82656152],
            [2.70687955],
            [3.86874928],
            [3.46117423],
            [3.        ],
            [2.44741977],
            [2.44741977],
            [2.81847013],
            [3.604023  ],
            [2.23387547],
            [3.59983881],
            [3.46007783],
            [3.31176749],
            [3.31055591],
            [1.99672481],
            [1.72962232],
            [3.31174976],
            [3.15601017],
            [2.44741977],
            [3.60073272],
            [4.57528521],
            [2.44948974],
            [2.        ],
            [2.82656152],
            [2.82842712],
            [3.31498372],
            [2.69902625]])
```

```
[10]: # LDA get topics for the reviews. works OK considering they're single sentences
      ↪:)
```

```
LDA = LatentDirichletAllocation(n_components = 20, random_state = 42)
LDA.fit(tfidf_matrix)

lda_components=LDA.components_
terms = tfidf.get_feature_names()
for index, component in enumerate(lda_components):
    top_terms_key=sorted(zip(terms, component), key = lambda t: t[1],
    ↪reverse=True)[:5]
    top_terms_list=list(dict(top_terms_key).keys())
    print("Topic "+str(index+1)+": ",top_terms_list)
```

```
Topic 1:  ['wells', 'the', 'good', 'an', 'inexperienced']
```

```

Topic 2: ['seems', 'even', 'movie', 'better', 'comedy']
Topic 3: ['demeo', 'material', 'needs', 'talent', 'without']
Topic 4: ['seems', 'even', 'movie', 'better', 'comedy']
Topic 5: ['wrong', 'hands', 'peploe', 'simply', 'unbearable']
Topic 6: ['seems', 'even', 'movie', 'better', 'comedy']
Topic 7: ['almost', 'fascinating', 'others', 'some', 'end']
Topic 8: ['constantly', 'implausibility', 'intelligence', 'maelstrom',
'observant']
Topic 9: ['cloying', 'drama', 'eventually', 'gripping', 'initially']
Topic 10: ['detached', 'hero', 'life', 'mood', 'oddly']
Topic 11: ['definitely', 'guilty', 'category', 'enjoyable', 'fire']
Topic 12: ['many', 'resurrection', 'one', 'entertaining', 'new']
Topic 13: ['apparently', 'bothered', 'check', 'featuring', 'fewer']
Topic 14: ['seems', 'even', 'movie', 'better', 'comedy']
Topic 15: ['miss', 'near', 'soggy', 'timid', 'even']
Topic 16: ['bore', 'casual', 'filmgoers', 'majority', 'probably']
Topic 17: ['fit', 'pieces', 'puzzle', 'whose', 'even']
Topic 18: ['asian', 'episode', 'gushy', 'perspective', 'time']
Topic 19: ['seems', 'even', 'movie', 'better', 'comedy']
Topic 20: ['acts', 'although', 'barbershop', 'blaxploitation', 'boasts']

```

```

[11]: colors = {'PERSON': 'linear-gradient(90deg, #aa9cfc, #fc9ce7)',
'DATE': 'linear-gradient(90deg, #34ebb4, #9c34eb)',
'ORG': 'linear-gradient(90deg, #eb3d34, #ebe834)',
'GPE': 'linear-gradient(90deg, #eb34d9, #bd0636)',
'TIME': 'linear-gradient(90deg, #9bbd06, #06bd89)',
'NORP': 'linear-gradient(90deg, #067dbd, #1206bd)',
'CARDINAL': 'linear-gradient(90deg, #06b4bd, #de0942)',
}
options = {'ents': ['PERSON', 'DATE', 'ORG', 'GPE', 'TIME', 'NORP',
↳ 'CARDINAL'], 'colors': colors}

```

```

[20]: # NER on the first 100 sentences with some funky colors I picked

for i in range(100):
    doc = nlp(sentence_samples_processed["sentence"][i])
    displacy.render(doc, style='ent', jupyter=True, options = options)

```

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

[illegible]





[illegible]

```
[13]: # WordClouds for the first 10 sentences
plt.figure(figsize=(20, 10))
fig, ax = plt.subplots(3, 3)
```

```
plt.axis("off")
for i in range(3):
    for j in range(3):
        text = sentence_samples_processed["sentence"][3*i + j]
        wordcloud = WordCloud(width=800, height=400, background_color="black").
        generate(text)
        ax[i,j].imshow(wordcloud, interpolation="bilinear")
plt.show()
```

<Figure size 2000x1000 with 0 Axes>



[14]: *# Translate it to spanish because - why not?*

```
for i in range(100):
    text = sentence_samples["sentence"][i]
    blob = TextBlob(text)
    translated_blob = blob.translate('en', 'es')
    print(translated_blob)
```

Aunque Barbershop cuenta con algunos de los actos más populares y modernos de hoy en el mundo de la televisión, la música y la comedia de pie, esta película extrañamente tiene la arrogancia desactualizada de una descarada incompleta de la comedia de sitios de los años 70 blaxploitation. Un rompecabezas cuyas piezas no encajan.

Algunos son fascinantes y otros no, y al final, es casi una buena película. ¿Ese Greengrass había sido un poco menos para la arena y mucho más por la inteligibilidad?

Lo bueno es muy, muy bueno ... el resto funciona de un poco impresionante a desesperadamente horrible.

'Butterfingers' es la palabra para la dirección de Jez Butterworth, que logra explotar incluso las sensibilidades más pequeñas del romance con su enfoque clamoroso.

Tenga cuidado, si está deprimido por algo antes de ver esta película, es posible que termine tratando de ahogarse en un lago después.

Una terrible adaptación de una obra que solo caminó por la delicada cuerda floja entre lastre y repugnantes.

En las manos equivocadas, es decir, Peploe, es simplemente insoportable

Un director inexperto, Mehta tiene mucho que aprender.

Un vehículo flojo de Eddie Murphy del que incluso él parece avergonzado de ser parte.

Tan confuso, repetitivo y irregular que dice mucho menos sobre la horrible realidad histórica que sobre el estilo característico del cineasta.

Un episodio chispeante de `` m \\* a \\* s \\* h '' solo esta vez desde una perspectiva asiática.

`` Buscando a Leonard '' parece sentarse en neutral, con la esperanza de que un viento rígido lo sople cuesta arriba o algo así.

Nada de cuatro o cinco risas suaves rodeadas por 86 minutos de comedia demasiado familiar y mal construida.

Definitivamente en la categoría de la película B de placer culpable, Reign of Fire es tan increíblemente inanimado que es risueño.

La buena pero implacablemente baja salida de poca brecha como CluisSess Does South Fork.

lo suficientemente entretenido, pero nada nuevo

... una resurrección demasiada.

Esta es una película sobre la naturaleza molesta y agotadora de la complacencia que sigue completamente satisfecha de seguir siendo la misma en todo momento.

Incluso cuando el héroe de la historia redescubre su pasión en la vida, el estado de ánimo permanece extrañamente separado.

DeMeo no está exento de talento; Solo necesita mejor material.

A pesar de presentar un guión acreditado a no menos de cinco escritores, aparentemente nadie aquí se molestó en comprobarlo dos veces.

Nadie involucrado, Save Dash, muestra la más mínima aptitud para la actuación, y el guión, acreditado al director Abdul Malik Abbott y Ernest 'Tron' Anderson, parece completamente improvisado.

Inicialmente apasionante, eventualmente empalagando el drama de POW.

Un tímido, empapado cerca de la señorita.

Funciona mejor en la concepción que en la ejecución ... termina pareciendo un poco demasiado inteligente.

Para la gran mayoría de los películas más casuales, probablemente será un aburrido.

La inteligencia observadora constantemente compite con la pretensión, y a veces una ingredacibilidad simple y extraña, en toda la vorágine.

Esta versión de H.G. Wells 'Time Machine fue dirigida por el bisnieto de H.G. Wells.

Deberían haber encontrado el bisnieto de Orson Welles.

All About Lily Chou Chou de Shunji Iwai es una película bellamente filmada, pero finalmente defectuosa sobre el crecimiento en Japón.

Con más desarrollo del personaje, este podría haber sido un misterioso thriller; Con mejores pagos, podría haber sido la película de monstruos de un hombre de pensamiento.

El debut como director de thriller para el escriba de tráfico Gaghan tiene todas las partes correctas, pero las piezas no se ajustan del todo.

... sería una pérdida total si no fuera por dos actuaciones de apoyo que tienen lugar en los bordes de la película.

No hay un solo momento de salto en el asiento y lo creas o no, Jason en realidad pasa a un segundo plano en su propia película a efectos especiales.

Goldbacher se basa en un sentido visual elegante y un talento para un ritmo fácil y seductor ... pero ella y su compañera de escritura Laurence Coriat no manejan una moneda narrativa igualmente asegurada.

Aunque Harris está afectando a veces, no puede superar la sensación de que la calabaza es un mero peón de la trama para dos directores con discapacidades mucho menos entrañables.

El documental es demasiado convencional, muchas cabezas parlantes aburridas, etc., para hacer justicia al tema.

La película en sí parece estar ejecutándose de hipertima a la inversa a medida que las partes verdaderamente divertidas se separan cada vez más.

Esta no es una película de Jackie Chan.

Es solo una película que tiene a Jackie Chan en ella.

Y eso hace toda la diferencia.

Demasiado inteligente a la mitad, la película de Howard es realmente una serie de momentos de rumbo, con todos los espacios intermedios llenos de fantasías, sueños, recuerdos y un fantástico tropo visual tras otro.

El problema con las películas sobre ángeles es que tienden a pasar a Hokum.

Un rumor de los ángeles no solo se desliza, está avalanchas en la confusión forzada.

No hay gran grito, nada nuevo para ver, cero emociones, demasiados flashbacks y un final entrecortado es una mala película.

No creo que esta película ama a las mujeres en absoluto.

Shankman ... y la guionista Karen Janszen Bungle a través de la narrativa como si fuera una serie de parábolas bíblicas y no una historia real.

Una comedia británica insignificante.

No puede convencer a la audiencia de que estos mocosos serán más que perdedores.

Flojo y sin inspiración, y poblados principalmente por personajes tan poco simpáticos que te queda con un sabor agrio en la boca.

Omita este turd y elija la nariz porque seguramente obtendrá más de la última experiencia.

¿Qué se puede decir sobre un actor calvo de 50 años que interpreta a un niño inocente tallado en un tronco?

Cinema de basura del trailer, tan poco cool lo único que falta es el ``

¡Gadzooks! ''

Su película es como una hermosa comida que no se calienta adecuadamente, de modo que termina un poco frío y relativamente sin sabor.

Al igual que el mundo de su película, Hartley creó un monstruo pero no sabía cómo manejarla.

No hay nuevas concepciones de la trama o cambios ambientales, solo cuerpos diferentes para que los objetos afilados lo atraviesen.

Necesita una cinematografía más impresionista y tomas estimulantes de punto de vista y menos tomas de "grandeza" de movimiento lento y ediciones de corte rápido que a menudo rayan del atletismo.

Al final, no hay mucho.

Un desperdicio de pureza intrépida en la artesanía de actuación.

La película es en última instancia tan inspiradora como una tarjeta Hallmark.

Se recomienda a cualquiera que no se dedique a las salpicaduras de alta tecnología que tome la advertencia literalmente e inicie sesión en algo más fácil de usar.

Las acciones de mala reputación y las exquisitas trampas se amortiguan por un script mediocre y actuaciones de calidad inferior.

Podrías confundirlo fácilmente con un trabajo incompleto en el progreso que se apresuró inexplicablemente a los megaplexes antes de su tiempo.

Los directores Harry Gantz y Joe Gantz han elegido un tema fascinante, pero las parejas que se exponen no son tan interesantes.

Otra entrada más en el género sentimental oh-ese-brits-brits que fue introducido por el Monty completo y aún se esfuerza por producir otro éxito.

Para aquellos para quienes el nombre Woody Allen fue una vez una garantía de algo fresco, a veces divertido y, por lo general, vale la pena, el final de Hollywood es una experiencia deprimente

Femme Fatale no ofrece nada más que un cebo y conmutador que está más allá de jugar justo con la audiencia.

¿Estamos lidiando con sueños, visiones o que nos digan lo que realmente sucedió como si fuera el tercer final de la pista?

Podría haber sido algo especial, pero dos cosas lo arrastran a la mediocridad: el malentendido de los ritmos de Marivaux del director Clare Peploe, y las limitaciones de Mira Sorvino como actriz clásica.

Neo-Noir esponjoso escondido detrás de las referencias de cine cursi.

Imagina a Susan Sontag enamorándose de Howard Stern.

Al igual que estar atrapado dentro de un enorme videojuego, donde las imágenes emocionantes y locas siguen pasando la cabeza y las mismas cosas ilógicas siguen sucediendo una y otra vez.

Debería haber valido la pena animar como un avance, pero está desprovisto de ingenio y humor.

Lo mejor de la película es su elenco agradable y divertido.

Estos tipos parecen geniales para tocar una cerveza, pero simplemente no son artistas divertidos.

Todo era tan superficial como el acento de bajo cebado forzado de Nueva Jersey Uma.

El director David Fincher y el escritor David Koepp Ca n't Surft It.

Finalmente, bajando de los estantes profundos de Miramax después de un par de intentos abortados, despertar en Reno presenta un argumento fuerte para dejar

que los perros dormidos se encuentren.

Una película que se siente como el episodio piloto de una nueva serie de TV de acción dirigida a adolescentes.

Una de las decepciones más altamente presentadas que he tenido la desgracia de ver en bastante tiempo.

La animación y los fondos son exuberantes e inventivos, pero regresan a Neverland nunca se las arregla para llevarnos a ese lugar evasivo y encantador donde suspendemos nuestra incredulidad.

El director Shekhar Kapur y los guionistas Michael Schiffer y Hossein Amini han tratado de modernizar y reconceptualizar las cosas, pero las barreras finalmente demuestran ser demasiado grandes.

El cine fuerte requiere un claro sentido de propósito, y en esa categoría tan importante, las cuatro plumas se quedan cortas.

La idea de ver esta película con una audiencia llena de adolescentes que se fijan en el humor de su cuerpo y el refuerzo de los estereotipos -lrb- de los cuales obtendrán mucho -rrb- me llena de repulsión.

Se convierte en la derivada, apoyándose en efectos CGI mal renderizados.

Cualquiera que reciba escalofríos de películas con agujeros de trama gigantes encontrará mucho para sacudirse y temblar en el anillo. '

Un gran pedo que viene de un director que comienza a parecerse al loco abuelo francés de alguien.

El script es un desastre, con mensajes empalagosos y personajes molestos.

Tanto el exceso y poco desnutrido ... la película puede llamarse un éxito sólido, aunque hay mucha evidencia aquí para indicar que Clooney podría tener mejor suerte la próxima vez.

Pasa, menos el humor retorcido y las imágenes sorprendentes que han hecho de Miike ... un héroe de culto.

Hollywood ha tomado un gran problema del vuelo imaginativo de Alfred Hitchcock a la importante pelusa de verano de Shyamalan.

El Maudlin de la película se centra en la enfermedad de la joven y sus ingenuos sueños juegan como el peor tipo de piloto de hilos de Hollywood.

Creo firmemente que una buena película de videojuegos aparecerá pronto.

También creo que Resident Evil no es todo.

Tiene el aire de una alondra sorprendentemente juvenil, una broma influenciada por el pop cuyos amuletos son inmediatamente aparentes y se usan con repetición.

La trama serpentea de agarre a caminar y de regreso.

Esto es cruel y misantrópico con solo afirmaciones débiles de surrealismo y comedia negra.

Ninguna cantidad de nostalgia por los días de gloria de Carvey puede disfrazar el hecho de que la nueva película es una película inferior para niños y que los considerables talentos de Carvey se desperdician en ella.

[15]: *# POS tagging for the first 10 sentences (more would get spammy)*

```
for i in range(10):
    words = word_tokenize(sentence_samples_processed["sentence"][i])
    print(sentence_samples["sentence"][i])
```

```
for j in sorted(list(set(nltk.pos_tag(words))): print(j)
print()
```

Although Barbershop boasts some of today 's hottest and hippest acts from the world of television , music and stand-up comedy , this movie strangely enough has the outdated swagger of a shameless ` 70s blaxploitation shuck-and-jive sitcom .

```
('acts', 'NNS')
('although', 'IN')
('barbershop', 'NN')
('blaxploitation', 'NN')
('boasts', 'NNS')
('comedy', 'NN')
('enough', 'RB')
('hippest', 'JJS')
('hottest', 'VBP')
('jive', 'JJ')
('movie', 'NN')
('music', 'NN')
('outdated', 'JJ')
('shameless', 'JJ')
('shuck', 'VBD')
('sitcom', 'NN')
('stand', 'VBP')
('strangely', 'RB')
('swagger', 'JJR')
('television', 'NN')
('today', 'NN')
('world', 'NN')
```

A puzzle whose pieces do not fit .

```
('a', 'DT')
('fit', 'VBP')
('pieces', 'NNS')
('puzzle', 'NN')
('whose', 'WP$')
```

Some are fascinating and others are not , and in the end , it is almost a good movie .

```
('almost', 'RB')
('end', 'VBP')
('fascinating', 'VBG')
('good', 'JJ')
('movie', 'NN')
('others', 'NNS')
('some', 'DT')
```

Would that Greengrass had gone a tad less for grit and a lot more for

intelligibility .  
( 'gone', 'VBN' )  
( 'greengrass', 'VB' )  
( 'grit', 'JJ' )  
( 'intelligibility', 'NN' )  
( 'less', 'RBR' )  
( 'lot', 'NN' )  
( 'tad', 'RB' )  
( 'would', 'MD' )

The good is very , very good ... The rest runs from mildly unimpressive to despairingly awful .

( 'awful', 'JJ' )  
( 'despairingly', 'RB' )  
( 'good', 'JJ' )  
( 'good', 'NN' )  
( 'mildly', 'RB' )  
( 'rest', 'NN' )  
( 'runs', 'VBZ' )  
( 'the', 'DT' )  
( 'unimpressive', 'JJ' )

` Butterfingered ' is the word for the big-fisted direction of Jez Butterworth , who manages to blast even the smallest sensitivities from the romance with his clamorous approach .

( 'approach', 'NN' )  
( 'big', 'JJ' )  
( 'blast', 'RB' )  
( 'butterfingered', 'VBN' )  
( 'butterworth', 'NN' )  
( 'clamorous', 'JJ' )  
( 'direction', 'NN' )  
( 'even', 'RB' )  
( 'fisted', 'VBN' )  
( 'jez', 'NN' )  
( 'manages', 'VBZ' )  
( 'romance', 'VBP' )  
( 'sensitivities', 'NNS' )  
( 'smallest', 'JJS' )  
( 'word', 'NN' )

Be forewarned , if you 're depressed about anything before watching this film , you may just end up trying to drown yourself in a lake afterwards .

( 'afterwards', 'NNS' )  
( 'anything', 'NN' )  
( 'be', 'VB' )  
( 'depressed', 'JJ' )  
( 'drown', 'JJ' )



```

('end', 'VB')
('film', 'NN')
('forewarned', 'VBN')
('lake', 'NN')
('may', 'MD')
('trying', 'VBG')
('watching', 'VBG')

```

a terrible adaptation of a play that only ever walked the delicate tightrope between farcical and loathsome .

```

('adaptation', 'NN')
('delicate', 'JJ')
('ever', 'RB')
('farcical', 'JJ')
('loathsome', 'NN')
('play', 'NN')
('terrible', 'JJ')
('tightrope', 'NN')
('walked', 'VBD')

```

In the wrong hands , i.e. Peploe 's , it 's simply unbearable

```

('e', 'VBP')
('hands', 'NNS')
('in', 'IN')
('peploe', 'NN')
('simply', 'RB')
('unbearable', 'JJ')
('wrong', 'JJ')

```

An inexperienced director , Mehta has much to learn .

```

('an', 'DT')
('director', 'NN')
('inexperienced', 'JJ')
('learn', 'NN')
('mehta', 'VBD')
('much', 'RB')

```

```

[16]: # Example on how to use gensim word2vec for similarity
text = ". ".join(sentence_samples_processed["sentence"][:250])
text = [nltk.word_tokenize(sent) for sent in nltk.sent_tokenize(text)]
model = gensim.models.Word2Vec(window=10, sg=1, hs=0, negative=10, alpha=0.03,
    ↪min_alpha=0.0007, seed=4)
model.build_vocab(text, progress_per=200)
model.train(text, total_examples=model.corpus_count, epochs=10, report_delay=1)
print("Similarity of 'film', 'movie': %s" % model.wv.similarity("film",
    ↪"movie"))

```

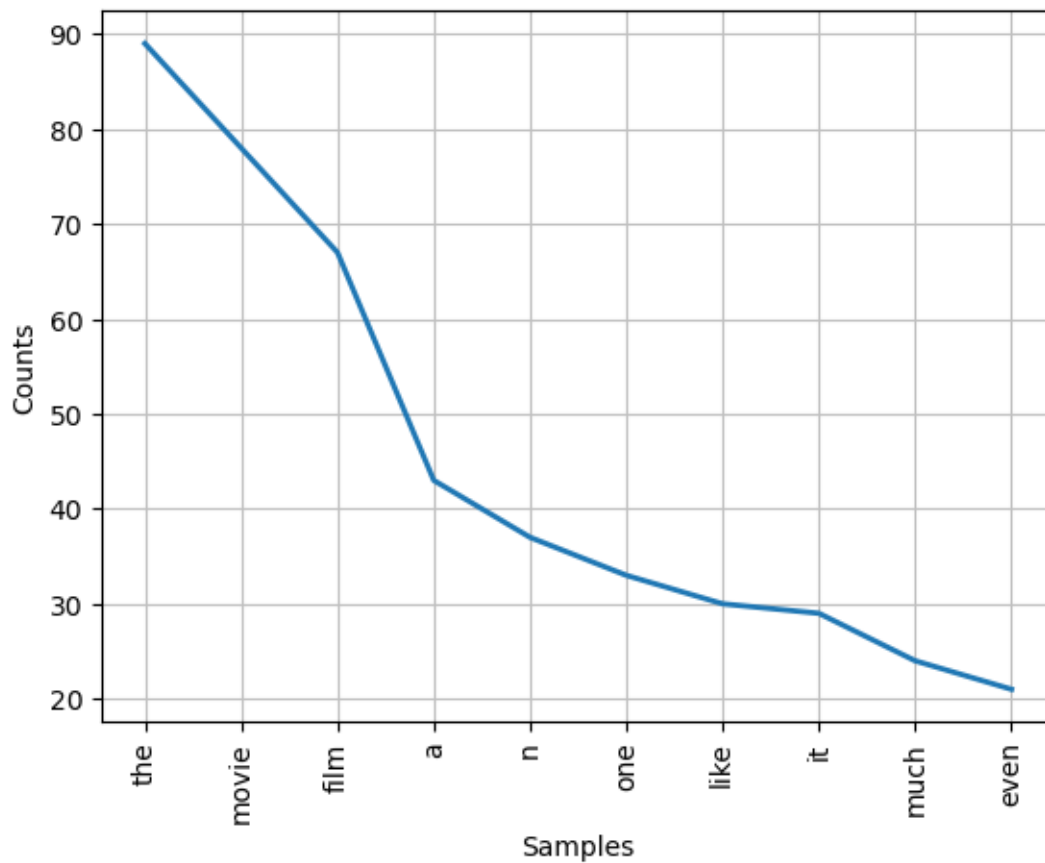
Similarity of 'film', 'movie': 0.9704475

```
[17]: # Most frequent words in our dataset using FreqDist from nltk
text = word_tokenize(" ".join(sentence_samples_processed["sentence"]))

fdist = FreqDist(text)
print("Most common words (with #):")
for i in fdist.most_common(10): print(i)
fdist.plot(10)
plt.show()
```

Most common words (with #):

```
('the', 89)
('movie', 78)
('film', 67)
('a', 43)
('n', 37)
('one', 33)
('like', 30)
('it', 29)
('much', 24)
('even', 21)
```



```
[18]: # Sentiment analysis using a BertTokenizer (pre-trained), on the first 50
      ↪ reviews
      model_name = "bert-base-uncased"
      tokenizer = BertTokenizer.from_pretrained(model_name)
      model = TFBertForSequenceClassification.from_pretrained(model_name)

      # Create a function to analyze sentiment
      def analyze_sentiment(sentence):
          # Tokenize the input text
          inputs = tokenizer(sentence, padding=True, truncation=True,
      ↪ return_tensors="tf")
          # Perform sentiment analysis
          output = model(inputs)
          predicted_class = tf.argmax(output.logits, axis=1).numpy()[0]
          return "Sentiment: " + ("Positive" if predicted_class == 1 else "Negative")
```

All PyTorch model weights were used when initializing TFBertForSequenceClassification.

Some weights or buffers of the TF 2.0 model TFBertForSequenceClassification were not initialized from the PyTorch model and are newly initialized:

```
['classifier.weight', 'classifier.bias']
```

You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.

```
[19]: for i in range(50):
      print("Review: " + sentence_samples["sentence"][i])
      print(analyze_sentiment(sentence_samples["sentence"][i]))
```

Review: Although Barbershop boasts some of today 's hottest and hippest acts from the world of television , music and stand-up comedy , this movie strangely enough has the outdated swagger of a shameless ` 70s blaxploitation shuck-and-jive sitcom .

Sentiment: Positive

Review: A puzzle whose pieces do not fit .

Sentiment: Positive

Review: Some are fascinating and others are not , and in the end , it is almost a good movie .

Sentiment: Positive

Review: Would that Greengrass had gone a tad less for grit and a lot more for intelligibility .

Sentiment: Positive

Review: The good is very , very good ... The rest runs from mildly unimpressive to despairingly awful .

Sentiment: Positive

Review: ` Butterfingered ' is the word for the big-fisted direction of Jez Butterworth , who manages to blast even the smallest sensitivities from the romance with his clamorous approach .

Sentiment: Positive

Review: Be forewarned , if you 're depressed about anything before watching this film , you may just end up trying to drown yourself in a lake afterwards .

Sentiment: Positive

Review: a terrible adaptation of a play that only ever walked the delicate tightrope between farcical and loathsome .

Sentiment: Positive

Review: In the wrong hands , i.e. Peploe 's , it 's simply unbearable

Sentiment: Positive

Review: An inexperienced director , Mehta has much to learn .

Sentiment: Positive

Review: A limp Eddie Murphy vehicle that even he seems embarrassed to be part of .

Sentiment: Positive

Review: So muddled , repetitive and ragged that it says far less about the horrifying historical reality than about the filmmaker 's characteristic style .

Sentiment: Positive

Review: A gushy episode of `` M \\* A \\* S \\* H '' only this time from an Asian perspective .

Sentiment: Positive

Review: `` Looking For Leonard '' just seems to kinda sit in neutral , hoping for a stiff wind to blow it uphill or something .

Sentiment: Positive

Review: Nothing more than four or five mild chuckles surrounded by 86 minutes of overly-familiar and poorly-constructed comedy .

Sentiment: Positive

Review: Definitely in the guilty pleasure B-movie category , Reign of Fire is so incredibly inane that it is laughingly enjoyable .

Sentiment: Positive

Review: Good-looking but relentlessly lowbrow outing plays like Clueless Does South Fork .

Sentiment: Positive

Review: entertaining enough , but nothing new

Sentiment: Positive

Review: ... one resurrection too many .

Sentiment: Positive

Review: This is a film about the irksome , tiresome nature of complacency that remains utterly satisfied to remain the same throughout .

Sentiment: Positive

Review: Even as the hero of the story rediscovers his passion in life , the mood remains oddly detached .

Sentiment: Positive

Review: DeMeo is not without talent ; he just needs better material .

Sentiment: Positive

Review: In spite of featuring a script credited to no fewer than five writers ,

apparently nobody here bothered to check it twice .  
 Sentiment: Positive  
 Review: No one involved , save Dash , shows the slightest aptitude for acting ,  
 and the script , credited to director Abdul Malik Abbott and Ernest ` Tron ' Anderson , seems entirely improvised .  
 Sentiment: Positive  
 Review: Initially gripping , eventually cloying POW drama .  
 Sentiment: Positive  
 Review: A timid , soggy near miss .  
 Sentiment: Positive  
 Review: Works better in the conception than it does in the execution ... winds up seeming just a little too clever .  
 Sentiment: Positive  
 Review: To the vast majority of more casual filmgoers , it will probably be a talky bore .  
 Sentiment: Positive  
 Review: Observant intelligence constantly vies with pretension -- and sometimes plain wacky implausibility -- throughout Maelstrom .  
 Sentiment: Positive  
 Review: This version of H.G. Wells ' Time Machine was directed by H.G. Wells ' great-grandson .  
 Sentiment: Positive  
 Review: They should have found Orson Welles ' great-grandson .  
 Sentiment: Positive  
 Review: Shunji Iwai 's All About Lily Chou Chou is a beautifully shot , but ultimately flawed film about growing up in Japan .  
 Sentiment: Positive  
 Review: With more character development this might have been an eerie thriller ; with better payoffs , it could have been a thinking man 's monster movie .  
 Sentiment: Positive  
 Review: Thriller directorial debut for Traffic scribe Gaghan has all the right parts , but the pieces do n't quite fit together .  
 Sentiment: Positive  
 Review: ... would be a total loss if not for two supporting performances taking place at the movie 's edges .  
 Sentiment: Positive  
 Review: There 's not a single jump-in-your-seat moment and believe it or not , Jason actually takes a backseat in his own film to special effects .  
 Sentiment: Positive  
 Review: Goldbacher draws on an elegant visual sense and a talent for easy , seductive pacing ... but she and writing partner Laurence Coriat do n't manage an equally assured narrative coinage .  
 Sentiment: Positive  
 Review: Though Harris is affecting at times , he can not overcome the sense that Pumpkin is a mere plot pawn for two directors with far less endearing disabilities .  
 Sentiment: Positive  
 Review: The documentary is much too conventional -- lots of boring talking heads

, etc. -- to do the subject matter justice .  
 Sentiment: Positive  
 Review: The movie itself appears to be running on hypertime in reverse as the truly funny bits get further and further apart .  
 Sentiment: Positive  
 Review: This is not a Jackie Chan movie .  
 Sentiment: Positive  
 Review: It 's just a movie that happens to have Jackie Chan in it .  
 Sentiment: Positive  
 Review: And that makes all the difference .  
 Sentiment: Positive  
 Review: Far too clever by half , Howard 's film is really a series of strung-together moments , with all the spaces in between filled with fantasies , daydreams , memories and one fantastic visual trope after another .  
 Sentiment: Positive  
 Review: The problem with movies about angels is they have a tendency to slip into hokum .  
 Sentiment: Positive  
 Review: A Rumor of Angels does n't just slip -- it avalanches into forced fuzziness .  
 Sentiment: Positive  
 Review: No big whoop , nothing new to see , zero thrills , too many flashbacks and a choppy ending make for a bad film .  
 Sentiment: Positive  
 Review: I do n't think this movie loves women at all .  
 Sentiment: Positive  
 Review: Shankman ... and screenwriter Karen Janszen bungle their way through the narrative as if it were a series of Bible parables and not an actual story .  
 Sentiment: Positive  
 Review: A negligible British comedy .  
 Sentiment: Positive

[ ]: