¿Cómo es tu personaje?

Un análisis de los personajes de tu serie favorita

Table of contents

01
Pensando
Corpus elegido y qué se busca analizar

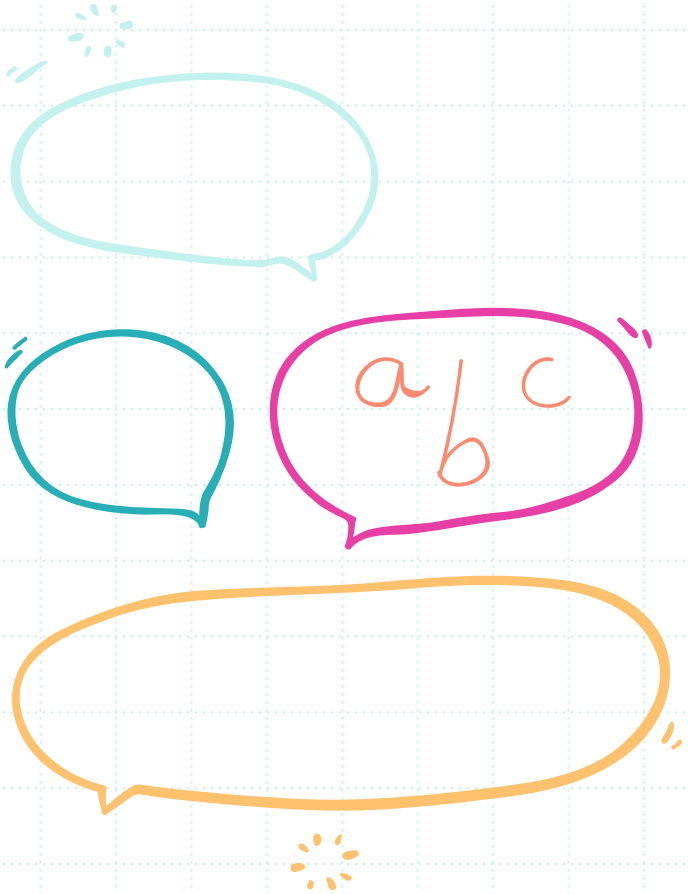
02
Scrapeando
Obtención del corpus y pasaje a texto

03
Tokenizando
Etiquetar correctamente mi corpus

04
Relacionando
Asociar los token a los personajes

05
Encodeando
Cómo adaptar el corpus para darselo a una red

06
Clasificando
Relacionar palabras con personajes

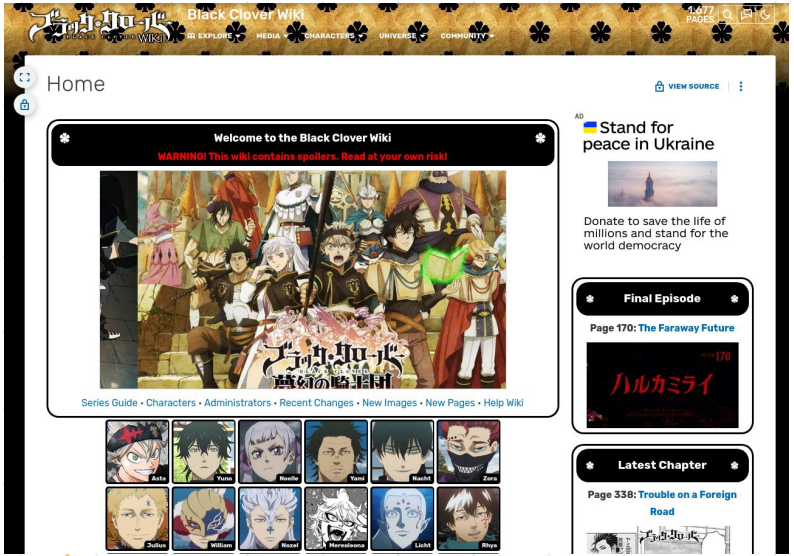


01

Pensando

Corpus elegido y qué se busca analizar

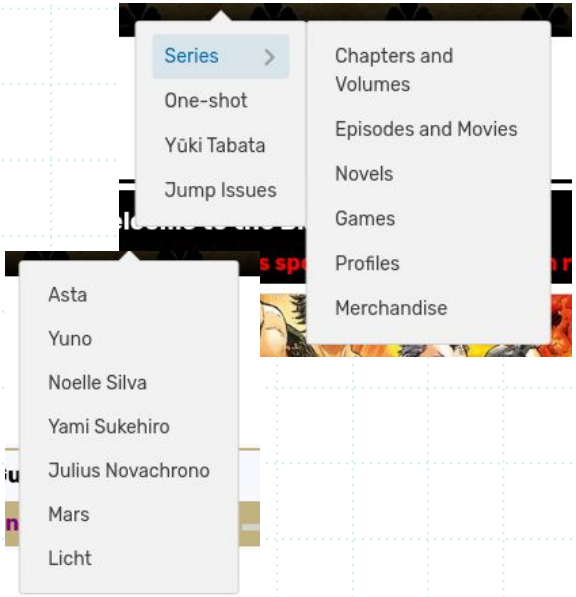
¿Qué es una Wiki?



Es como una Wikiapedia de un tema específico.

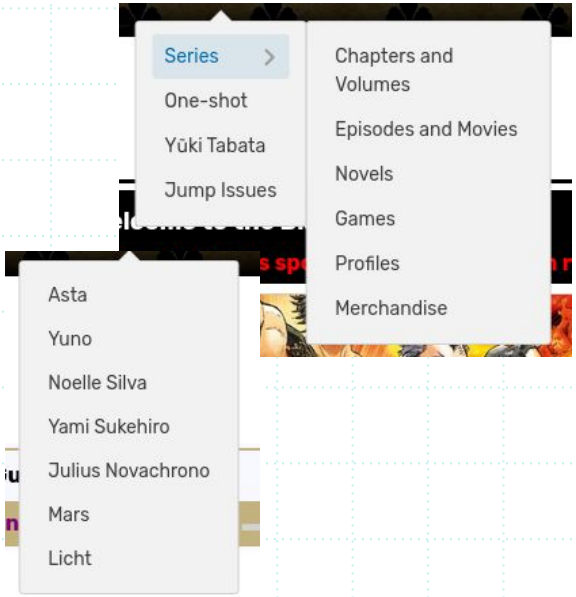
- Contenido **escrito por usuarios** y cualquier usuario puede realizar cambios
- Cualquiera puede decidir abrir una nueva (puede haber más de una que trate un tema)
- Es muy común que se **centren en algún mundo ficcional** (una serie, universo de series, peliculas, sagas de libros, etc).

¿Qué tiene una Wiki?



- Descripciones detalladas de episodios
- Descripciones detalladas de personajes
- Libros, Videojuegos, Merchandise, etc.
- Cualquier cosa que el Fandom haya decidido escribir

¿Qué tiene una Wiki?



- Descripciones detalladas de episodios
- ~~Descripciones detalladas de personajes~~
- ~~Libros, Videojuegos, Merchandise, etc.~~
- ~~Cualquier cosa que el Fandom haya decidido escribir~~

¿Descripciones detalladas?

in: Witches' Forest Arc Episodes

English

Episode 61

VIEW SOURCE

The Promised World 「約束の世界 *Yakusoku no Sekai*」 is the 61st Page of Yūki Tabata's *Black Clover*.

Contents [hide]

1. Summary
2. Fights
3. Events
4. Magic and Spells used
5. Petit Clover
6. References
7. Navigation

Episode Information

Japanese	約束の世界
Rōmaji	*Yakusoku no Sekai*
Funimation	The Promised World
Airdate	December 4, 2018 December 23, 2018 (SimulDub) March 16, 2019 (Toonami)
Arc	Witches' Forest Arc
Chapter(s)	Chapter 91 Chapter 92 Chapter 93

Opening	**Ending**
Gamushara	Tenjou Tenge

Summary

Asta, Finral Roulacase, Noelle Silva, and Vanessa Enoteca have Fana surrounded. Above, Fanzell Kruger and Mars lead Ladros towards Asta but realize that he may have his hands full with Fana. Fanzell and Mars fly overhead just as the Salamander sends a fireball towards Asta. Fanzell tells Asta to deflect towards him. Asta does so and Fanzell aims the fireball towards Ladros.[1] It engulfs him and he burns up before falling towards the forest. Fanzell and Mars land in front of Asta. Asta asks Fanzell why he is with Mars since he is the enemy. Fanzell says he will explain everything later but for now he asks Asta to trust Mars. Asta reluctantly agrees. They turn towards Fana and Mars realizes that her flames are similar to his. He gasps as he recognizes Fana.[2]

Mars remembers when he was a child and how Fana used to cheer him up after tough training sessions. He also remembers how he and the other children were told to kill each other and that the winner would be the kingdom's mage warrior. Fana tried to kill him but Mars struck first. After, he was mourning when Fana reaches up and heals him. She told him to go see the world outside before she died. In the present, Mars yells out Fana's name and tries to go to her but is stopped by Asta. Fanzell recognizes her as well but says that he had heard she had died after having been experimented on by Moris. He wonders about the third eye she has on her forehead and believes that the Eye of the Midnight Sun is controlling her through it. Fana floats in the air and tells everyone that she hates them before she calls upon self-destruct magic.[3]

Fana sends multiple blasts towards them. Asta deflects some, Noelle creates a water barrier, but Mars lets the blasts hit him. Fanzell tells him to block it with his magic but Mars says that he has no right to because he killed Fana. Asta hits him and tells him that now is not the time for that. He tells Mars that if Fana is so important to him that now is not the time for him to die. He tells him to live so that he can save her. Fanzell has faith that Mars will be able to save her. Vanessa believes that her Thread Magic and Finral's Spatial Magic are no match for Fana. Noelle concludes that Asta needs to be the one to stop her.

¿Descripciones detalladas?

"Asta, Finral Roulacase, Noelle Silva, and Vanessa Enoteca have Fana surrounded. Above, Fanzell Kruger and Mars lead Ladros towards Asta but realize that he may have his hands full with Fana. Fanzell and Mars fly overhead just as the Salamander sends a fireball towards Asta. Fanzell tells Asta to deflect towards him. Asta does so and Fanzell aims the fireball towards Ladros. It engulfs him and he burns up before falling towards the forest. Fanzell and Mars land in front of Asta. Asta asks Fanzell why he is with Mars since he is the enemy. Fanzell says he will explain everything later but for now he asks Asta to trust Mars. Asta reluctantly agrees. They turn towards Fana and Mars realizes that her flames are similar to his. He gasps as he recognizes Fana.\nMars remembers when he was a child and how Fana used to cheer him up after tough training sessions. He also remembers how he and the other children were told to kill each other and that the winner would be the kingdom's mage warrior. Fana tried to kill him but Mars struck first. After, he was mourning when Fana reaches up and heals him. She told him to go see the world outside before she died. In the present, Mars yells out Fana's name and tries to go to her but is stopped by Asta. Fanzell recognizes her as well but says that he had heard she had died after having been experimented on by Moris. He wonders about the third eye she has on her forehead and believes that the Eye of the Midnight Sun is controlling her through it. Fana floats in the air and tells everyone that she hates them before she calls upon self-destruct magic.\nFana sends multiple blasts towards them. Asta deflects some, Noelle creates a water barrier, but Mars lets the blasts hit him. Fanzell tells him to block it with his magic but Mars says that he has no right to because he killed Fana. Asta hits him and tells him that now is not the time for that. He tells Mars that if Fana is so important to him that now is not the time for him to die. He tells him to live so that he can save her. Fanzell has faith that Mars will be able to save her. Vanessa believes that her Thread Magic and Finral's Spatial Magic are no match for Fana. Noelle concludes that Asta needs to be the one to stop her.\nAsta and Mars rush forward together. When they get closer, the heat from Fana's magic burns them, so Mars cloaks them in Phoenix Feathers Robe. As Asta walks forward, Mars is amazed by him. He has fought many other mages but he has never met anyone as stubborn as Asta. He believes that Asta is a true knight because he is always trying to protect people. Asta notices that Mars is not the same as when he fought him. Fanzell tells the others that they should evacuate because Fana could blow at any minute. Vanessa, Noelle, and Finral refuse to leave their friend behind. Fana tells them that it is over but Asta cuts through her magic with his sword.\nAsta thinks that it is over but Fana says it is not. Asta turns to face her but Mars runs past him and pulls Fana into a hug. Memories flash before Fana's eyes. Mars tells her that they must see the outside world together. The fire surrounding them surges but then disappears completely. Mars pulls back and Fana tells him that she promises to see the world with him. They embrace as Mars cries. Everyone else is relieved that it is over.\n"

¿Qué propongo?

A partir de las **descripciones de los episodios de una serie**, escritas por usuarios, **describir a los personajes de la serie**.

- De la Wikia solo se obtienen las descripciones de los episodios
- En una primera instancia lo que se busca es obtener calificativos de cada personaje

Summary

Asta, Finral Roulacase, Noelle Silva, and Vanessa Enoteca have Fana surrounded. Above, Fanzell Kruger and Mars lead Ladros towards Asta but realize that he may have his hands full with Fana. Fanzell and Mars fly overhead just as the Salamander sends a fireball towards Asta. Fanzell tells Asta to deflect towards him. Asta does so and Fanzell aims the fireball towards Ladros.[1] It engulfs him and he burns up before falling towards the forest. Fanzell and Mars land in front of Asta. Asta asks Fanzell why he is with Mars since he is the enemy. Fanzell says he will explain everything later but for now he asks Asta to trust Mars. Asta reluctantly agrees. They turn towards Fana and Mars realizes that her flames are similar to his. He gasps as he recognizes Fana.[2]

Mars remembers when he was a child and how Fana used to cheer him up after tough training sessions. He also remembers how he and the other children were told to kill each other and that the winner would be the kingdom's mage warrior. Fana tried to kill him but Mars struck first. After, he was mourning when Fana reaches up and heals him. She told him to go see the world outside before she died. In the present, Mars yells out Fana's name and tries to go to her but is stopped by Asta. Fanzell recognizes her as well but says that he had heard she had died after having been experimented on by Moris. He wonders about the third eye she has on her forehead and believes that the Eye of the Midnight Sun is controlling her through it. Fana floats in the air and tells everyone that she hates them before she calls upon self-destruct magic.[3]

Fana sends multiple blasts towards them. Asta deflects some, Noelle creates a water barrier, but Mars lets the

"Perseverante"

"Impulsivo"

"Heroico"

"Amigable"

Pero... ¿y el corpus?

170 episodios (al momento)

106 personajes (aprox.)

366 episodios (al momento)

534 personajes (según su wiki)

Pero... ¿y el corpus?

- Personajes simples (nada de actitudes cambiantes)
- Fácilmente describibles
- Muchos episodios
- Si bien se centran más en ciertos personajes, el elenco principal es grande
- Muchos personajes secundarios reciben bastante desarrollo
- Conocimiento personal de las series

La pregunta en concreto:

¿Cómo puedo **describir a los personajes** de la serie **a partir de sus acciones**?

Algunos problemas...

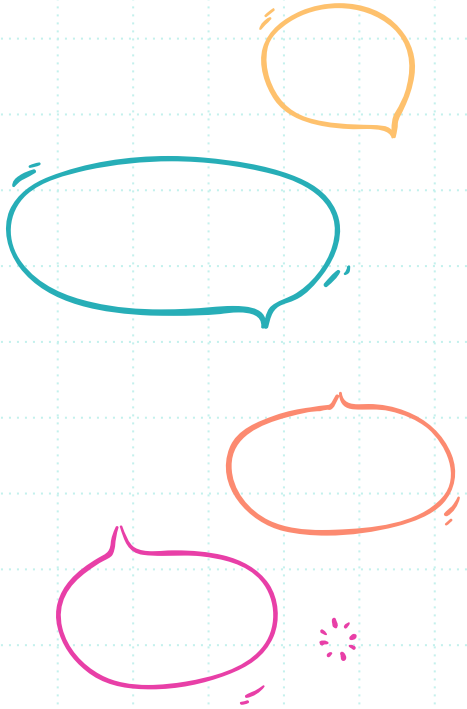
- Puede que se encuentren **typos** en el corpus *(hay algunos nombres difíciles que puede que se romanicen diferente según quién escriba)*
- Muchos usuarios usan **lenguaje coloquial e informal** (*phrasal verbs, idioms, etc*)
- No es un texto escrito por académicos por lo que hay **sobreuso de ciertas palabras** (*and, then, so, however, etc*).
- Los episodios son **relatados** por lo que se habla en tercera persona de los personajes (*Asta took his sword and blocked the attack*).

Todos estos problemas van a tener que ser tratados de alguna forma para obtener resultados deseables

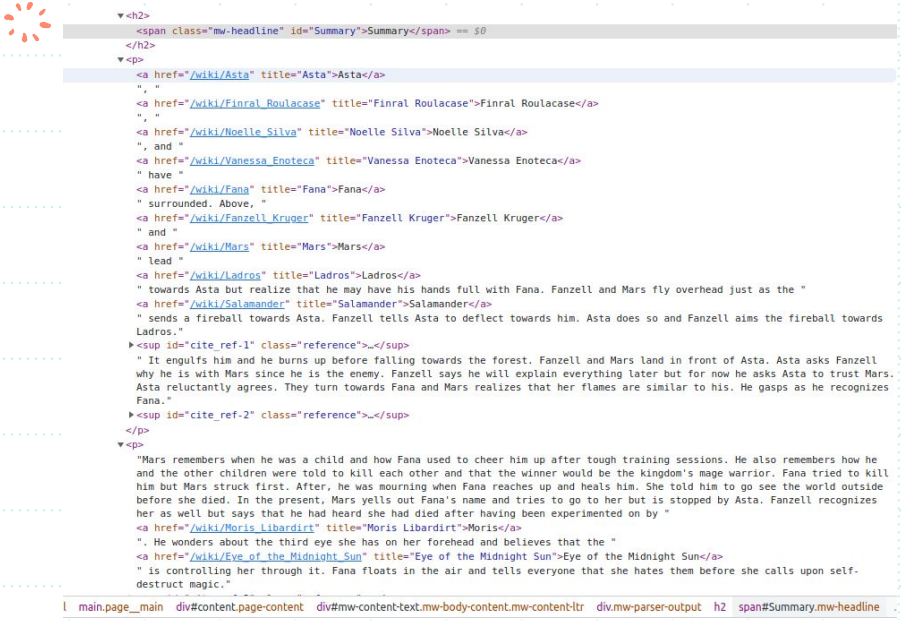
02

Scrapeando

Obtención del corpus y pasaje a texto



Las descripciones



- Webscrapping
- Los textos contienen tags de diferente índole
  - a => referencia a otra entrada de la wikia, es importante la info contenida
  - p => parrafo, este texto tiene parte de la descripción
  - sup => referencia externa, no tiene nada relevante

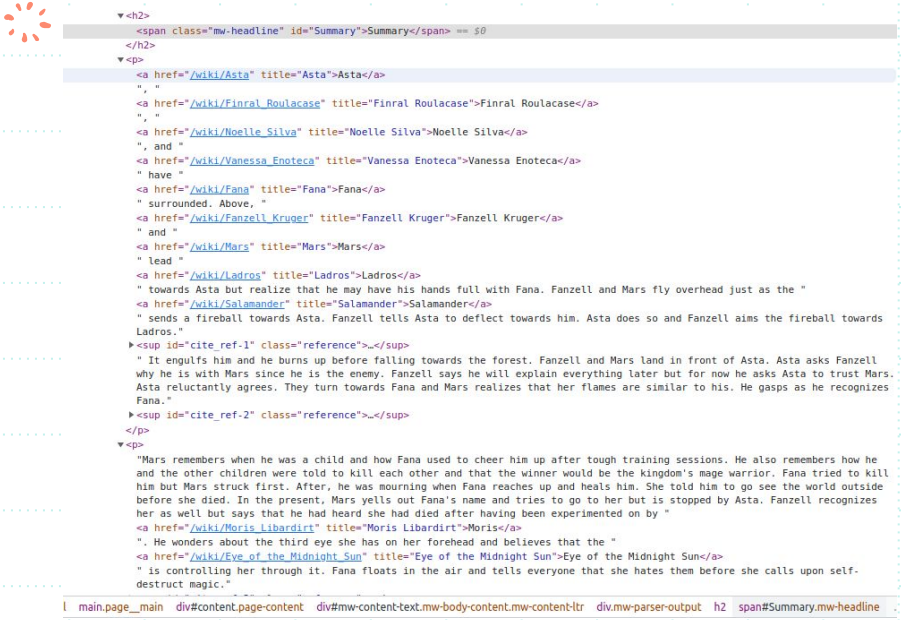
Las descripciones

```
<h2>
  <span class="mw-headline" id="Summary">Summary</span> == $0
</h2>
<p>
  <a href="/wiki/Asta" title="Asta">Asta</a>
  ", "
  <a href="/wiki/Finral_Roulacase" title="Finral Roulacase">Finral Roulacase</a>
  ", "
  <a href="/wiki/Noelle_Silva" title="Noelle Silva">Noelle Silva</a>
  ", and "
  <a href="/wiki/Vanessa_Enoteca" title="Vanessa Enoteca">Vanessa Enoteca</a>
  " have "
  <a href="/wiki/Fana" title="Fana">Fana</a>
  " surrounded. Above, "
  <a href="/wiki/Fanzell_Kruger" title="Fanzell Kruger">Fanzell Kruger</a>
  " and "
  <a href="/wiki/Mars" title="Mars">Mars</a>
  " lead "
  <a href="/wiki/Ladros" title="Ladros">Ladros</a>
  " towards Asta but realize that he may have his hands full with Fana. Fanzell and Mars fly overhead just as the "
  <a href="/wiki/Salamander" title="Salamander">Salamander</a>
  " sends a fireball towards Asta. Fanzell tells Asta to deflect towards him. Asta does so and Fanzell aims the fireball towards Ladros."
  <sup id="cite_ref-1" class="reference">…</sup>
  " It engulfs him and he burns up before falling towards the forest. Fanzell and Mars land in front of Asta. Asta asks Fanzell why he is with Mars since he is the enemy. Fanzell says he will explain everything later but for now he asks Asta to trust Mars. Asta reluctantly agrees. They turn towards Fana and Mars realizes that her flames are similar to his. He gasps as he recognizes Fana."
  <sup id="cite_ref-2" class="reference">…</sup>
</p>
<p>
  "Mars remembers when he was a child and how Fana used to cheer him up after tough training sessions. He also remembers how he and the other children were told to kill each other and that the winner would be the kingdom's mage warrior. Fana tried to kill him but Mars struck first. After, he was mourning when Fana reaches up and heals him. She told him to go see the world outside before she died. In the present, Mars yells out Fana's name and tries to go to her but is stopped by Asta. Fanzell recognizes her as well but says that he had heard she had died after having been experimented on by "
  <a href="/wiki/Moris_Libardirt" title="Moris Libardirt">Moris</a>
  ". He wonders about the third eye she has on her forehead and believes that the "
  <a href="/wiki/Eye_of_the_Midnight_Sun" title="Eye of the Midnight Sun">Eye of the Midnight Sun</a>
  " is controlling her through it. Fana floats in the air and tells everyone that she hates them before she calls upon self-destruct magic."

main.page__main  div#content.page-content  div#mw-content-text.mw-body-content.mw-content-ltr  div.mw-parser-output  h2  span#Summary.mw-headline
```

- Los textos contienen tags de diferente índole
  - i,b => solo le dan formato al texto, hay que tomar lo que contienen
  - figure => imagen, no es relevante
  - br => agregar un salto de línea extra, se debe descartar

Las descripciones



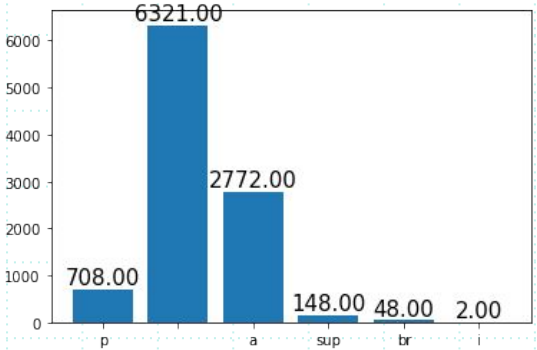
- Entre la sección de Summary y la próxima sección no hay diferencia de anidación (están todas al mismo nivel)
- Black Clover es sencillo obtener por episodio
  - La url es /Episode_61
- Bleach es difícil obtener por episodio
  - La url es /Nombre_del_ep
  - Se tuvo que scrapear la referencia al siguiente y **avanzar referencia por referencia**

Las descripciones

```
eps_summary, soup_tag_count = getEpsSummary(1, 161)
print(soup_tag_count)
```

1m 0.6s

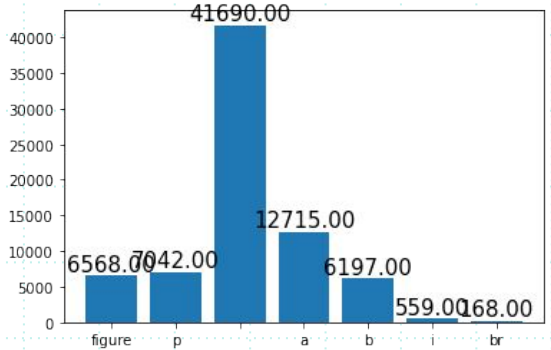
```
{'p': 708, None: 6321, 'a': 2772, 'sup': 148, 'br': 48, 'i': 2}
```




```
eps_summary, soup_tag_count = getEpsSummary("/wiki/The_Day_I_Became_a_Shinigami", 366)
print(soup_tag_count)
```

57.1s

```
{'figure': 6568, 'p': 7042, None: 41690, 'a': 12715, 'b': 6197, 'i': 559, 'br': 168}
```



Las descripciones

```
eps_summary, soup_tag_count = getEpsSummary(1, 161)
print(soup_tag_count)
```

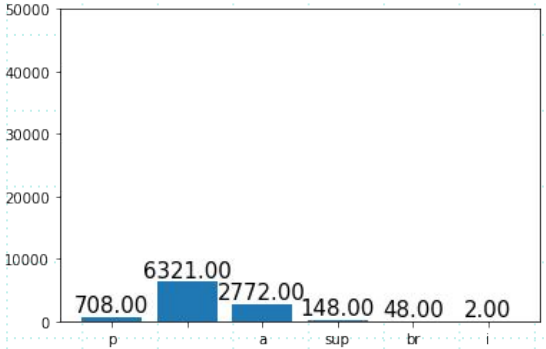
1m 0.6s

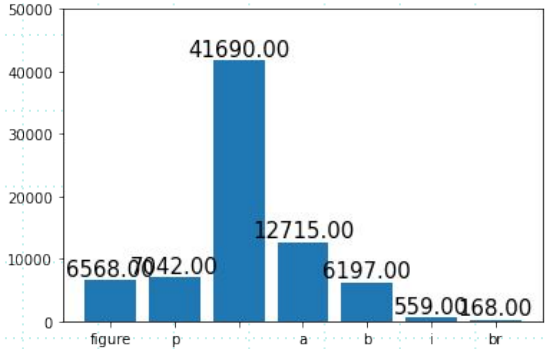
```
{'p': 708, None: 6321, 'a': 2772, 'sup': 148, 'br': 48, 'i': 2}
```

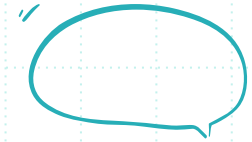
```
eps_summary, soup_tag_count = getEpsSummary("/wiki/The_Day_I_Became_a_Shinigami", 366)
print(soup_tag_count)
```

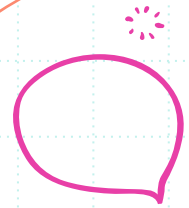
57.1s

```
{'figure': 6568, 'p': 7042, None: 41690, 'a': 12715, 'b': 6197, 'i': 559, 'br': 168}
```





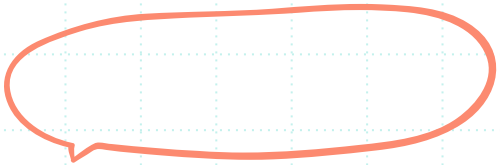




03

Tokenizando

Etiquetar correctamente mi corpus



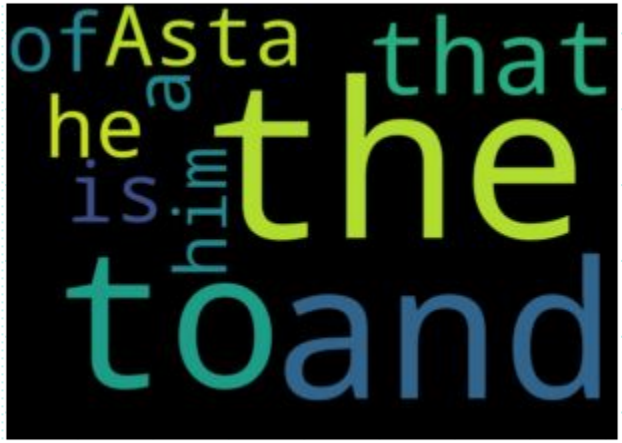
Expresiones Regulares

['In', 'Hage', 'a', 'priest', 'finds', 'two', 'babies', 'abandoned', 'outside', 'his', 'church', 'He', 'takes', 'them', 'inside', 'and', 'discovers', 'their', 'names', 'to', 'be', 'Yuno', 'and', 'Asta', 'Fifteen', 'years', 'later', 'Asta', 'proposes', 'to', 'Sister', 'Lily', 'who', 'refuses', 'repeatedly', 'Yuno', 'and', 'the', 'other', 'orphans', 'criticize', 'Asta', 'and', 'point', 'out', 'his', 'lack', 'of', 'magic', 'Asta', 'tries', 'to', 'show', 'off', 'his', 'skills', 'but', 'Yuno', 'outshines', 'him', 'with', 'his', 'magic', 'Later', 'at', 'the', 'Grimoire', 'Acceptance', 'Ceremony', 'a', 'pair', 'of', 'nobles', 'criticize', 'the', 'commoners', 'there', 'Despite', 'Asta', 'not', 'receiving', 'a', 'grimoire', 'while', 'Yuno', 'attains', 'a', 'four-leaf', 'clover', 'one', 'Asta', 'challenges', 'Yuno', 'to', 'the', 'title', 'of', 'Magic', 'Emperor', 'but', 'Yuno', 'ignores', 'him', 'After', 'the', 'ceremony', 'the', 'two', 'nobles', 'ambush', 'Yuno', 'outside', 'the', 'tower', 'but', 'Yuno', 'overpowers', 'them', 'All', 'three', 'are', 'then', 'ambushed', 'by', 'Revchi', 'Salik', 'a', 'former', 'Magic', 'Knight', 'who', 'plans', 'to', 'steal', "Yuno's", 'grimoire', 'and', 'sell', 'it', 'Asta', 'comes', 'to', "Yuno's", 'rescue', 'but', 'fails', 'As', 'Revchi', 'breaks', "Asta's", 'spirit', 'Yuno', 'calls', 'Asta', 'his', 'rival', 'Motivated', 'by', 'this', 'acknowledgment', 'Asta', 'decides', 'to', 'keep', 'fighting', 'which', 'summons', 'a', 'grimoire', 'from', 'the', 'tower', 'Revchi', 'notices', 'that', 'it', 'has', 'a', 'five-leaf', 'clover', 'and', 'when', 'a', 'large', 'sword', 'appears', 'from', 'the', 'grimoire', 'Asta', 'grabs', 'it', 'and', 'the', 'shadow', 'of', 'a', 'devil', 'appears', 'behind', 'him']

Palabras más mencionadas

the	4098
to	3128
and	3093
that	2100
he	1270
is	1254
Asta	1248
of	1157
a	1144
him	934

dtype: int64



La gran mayoría son **Stopwords**

Podemos notar que la mayoría de las palabras se **refieren al protagonista**

Palabras más mencionadas

Asta	1248
Magic	571
it	534
for	528
not	504
tells	477
as	455
are	454
Yami	437
magic	391

dtype: int64

The **Magic Knights** 「魔法騎士 *Mahō Kishi*」 are members of the Clover Kingdom's nine squads of mages who serve as protectors of the country. They are led by the Magic Emperor.[1]

Tiene sentido ver “magia” tan arriba ya que es el **concepto central**

Es interesante que hay “Magic” (propio) y “magic” (sustantivo / adjetivo)

Personajes más mencionados

Asta	1248
Yami	437
Noelle	388
Yuno	379
Julius	245
Finral	197
Licht	194
Vanessa	160
Magna	157
Gauche	134
Fanzell	134

dtype: int64

Nos da una idea de cuanta info tenemos de cada personaje

Yuno y Asta son los que comienzan la historia, sorprendentemente Yami y Noelle están más arriba que Yuno

Yami es el segundo más mencionado pese a que no sea del grupo protagónico.

Pronombres

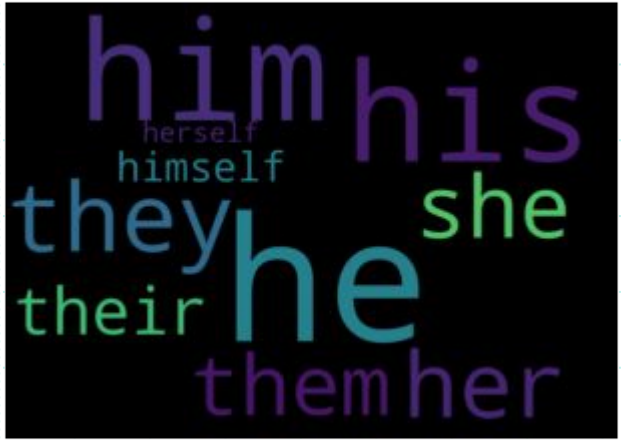
```
he     1536
dtype: int64

she     723
dtype: int64
```

```
him         934
his         915
himself      89
dtype: int64

her         808
herself      33
dtype: int64

they          880
them          680
their         376
themselves     25
theirs          1
dtype: int64
```



Pocos personajes femeninos. Se centra principalmente en los personajes masculinos.

Hay un personaje femenino (Noelle) que acapara la gran mayoría de las referencias a personajes femeninos

Palabras más mencionadas

```
magic      962
kingdom    185
knights    185
attack     175
one        167
mana       118
time       116
help       114
sword      113
bulls      112
dtype: int64
```

Tienen sentido dada la idea de “Un mundo mágico medieval”

NLTK: Pipeline

Tokenizar los documentos

```
['A city is being attacked by a giant demon.',
 'The First Magic Emperor then appears and defea
 'Over the following generations, the Clover Kin
 'The duty of all is to protect the kingdom, as
 'Asta finishes his thousand swings with the swo
 'Yuno has been training too, chopping wood and
 'Both of them want to apply to the Magic Knight
```

Tokenizar las oraciones

```
['However',
 ',',
 'Asta',
 'yells',
 'at',
 'him',
 'not',
```

Pasar a minúscula los tokens obtenidos, salvo aquellos que correspondan a Personajes

```
['however',
 ',',
 'Asta',
 'yell',
 'at',
 'him',
 'not',
```

Si no es un personaje => pasa a minúscula

Si es personaje => queda igual (primera mayúscula, resto minúscula)

Usar el tagger de NLTK para obtener la clase de palabra

```
[('however', 'RB'),
 (',', ','),
 ('Asta', 'NNP'),
 ('yell', 'VBP'),
 ('at', 'IN'),
 ('him', 'PRP'),
 ('not', 'RB'),
```

Quitar Stopwords

```
however
Asta
yell
give
declaring
prove
anyone
```

Lematizar

Minúsculas vs Vainilla

Minúsculas

```
{'kahono',
 'kaiser',
 'kaito',
 'kiato',
 'kid',
 'kingdom',
 'kira',
 'kirsch',
 'klaus',
 'magna',
 'mana',
 'marron',
 'xerx',
 'xiii',
 'yami',
 'yuno',
 'zara',
 'zenon',
 'zora'}
```

Palabras que son nombres propios, no las puede identificar

```
{'-50',
 'A',
 'Abari',
 'Above',
 'Abra',
 'Acceptance',
 'Acier',
 'Adlai',
 'Ado',
 'Afterward',
 'Agrippa',
 'Aguirre',
 'Air',
 'Alecdora',
 'Alfred',
 'Alive',
 'Anastasis',
 'Angel',
 'Angered',
 'Angry',
 'Anna',
 'Anti',
 'Anti-birds',
 'Aphro',
 'Aqua',
...
 'Zora',
 'Zuta',
 'kill',
 'magic',
 'undergo'}
```

Vainilla

Palabras que NO son nombres propios, las identifica como tal

Minúsculas salvo nombres vs Vanilla


```
{'Acier',
 'Asta',
 'Charlotte',
 'Charmy',
 'Damnatio',
 'Dorothy',
 'Fana',
 'Fanzell',
 'Finral',
 'Fuegoleon',
 'Gauche',
 'Gordon',
 'Grey',
 'Henry',
 'Jack',
 'Julius',
 'Kahono',
 'Kiato',
 'Kirsch',
 'Klaus',
 'Langris',
 'Lemiel',
 'Leopold',
 'Licht',
 'Lily',
...
 'kruger',
 'magic',
 'marron',
 'undergo',
 'xiii'}
```


Minúsculas salvo nombres

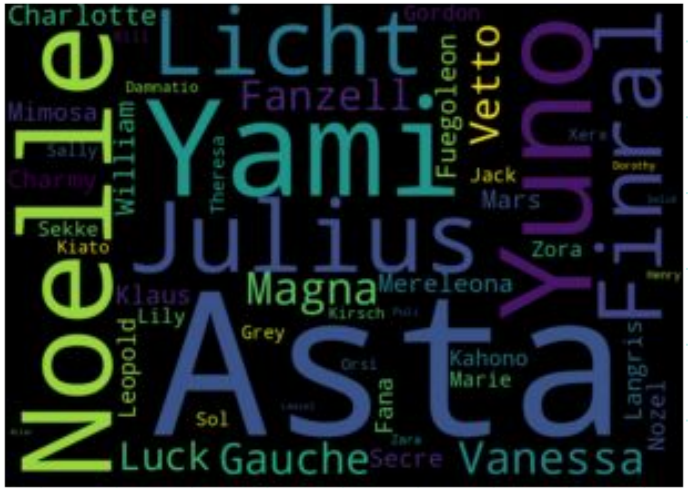
Logra identificar nombres correctamente con cierto margen de error

Vainilla

```
{'-50',
 'A',
 'Abari',
 'Above',
 'Abra',
 'Acceptance',
 'Acier',
 'Adlai',
 'Ado',
 'Afterward',
 'Agrippa',
 'Aguirre',
 'Air',
 'Alecdora',
 'Alfred',
 'Alive',
 'Anastasis',
 'Angel',
 'Angered',
 'Angry',
 'Anna',
 'Anti',
 'Anti-birds',
 'Aphro',
 'Aqua',
...
 'Zora',
 'Zuta',
 'kill',
 'magic',
 'undergo'}
```

Personajes más mencionados

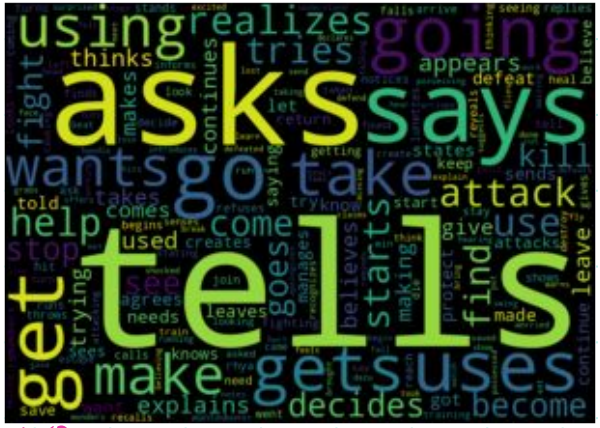
Asta	1322
Yami	448
Yuno	407
Noelle	399
Julius	257
Licht	220
Finral	207
Magna	161
Vanessa	161
Gauche	142
Fanzell	137
Vetto	127
Luck	117
Charlotte	112
Fuegoleon	109
William	108
Klaus	98
Mimosa	95
Charmy	93
Mereleona	92
Mars	90
Secre	87
Langris	83
Leopold	81
Fana	79
...	
Dorothy	10
Solid	9
Lemiel	9
Acier	6
dtype: int64	



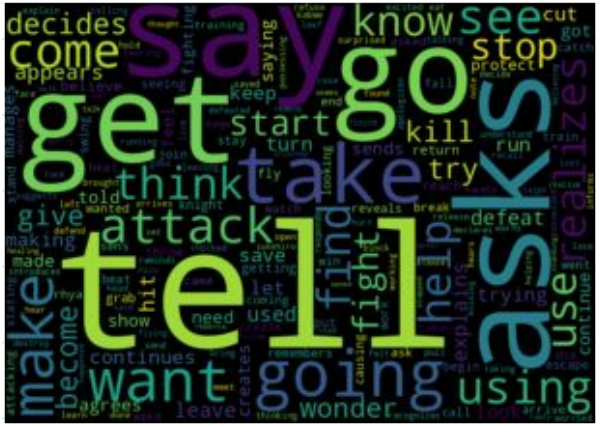
Asta	1248
Yami	437
Noelle	388
Yuno	379
Julius	245
Finral	197
Licht	194
Vanessa	160
Magna	157
Gauche	134
Fanzell	134
dtype: int64	

Verbos más mencionados

Sin Lematizar



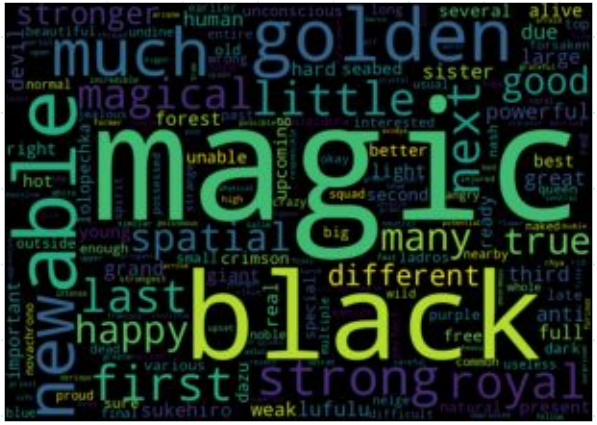
Lematizado



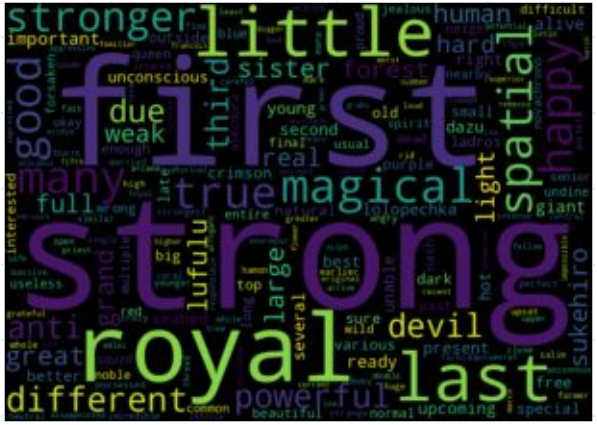
Filtrado

Adjetivos más mencionados

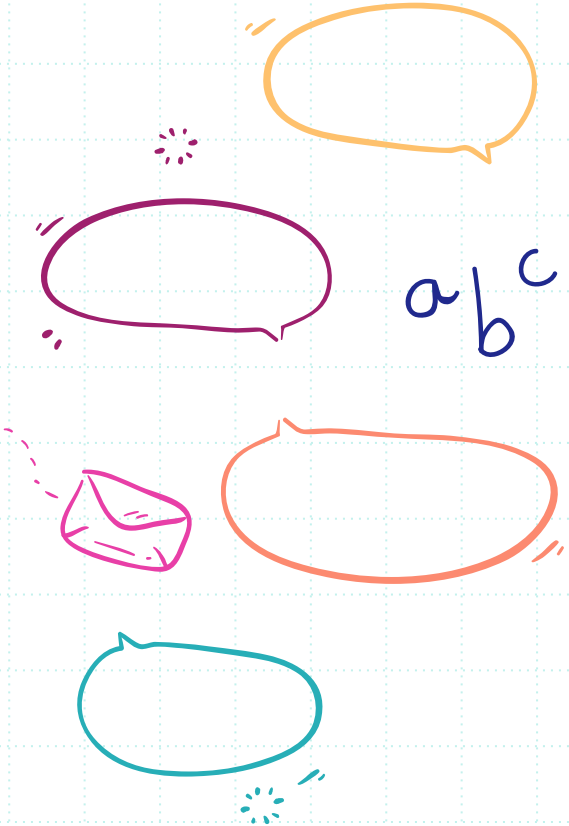
Sin Filtrar



Sin Lematizar



Lematizado



04

Relacionando

Asociar los token a los personajes

Idea

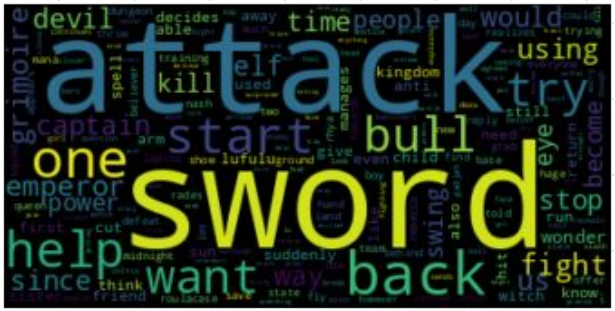
Vincular a cada personaje con las partes del corpus donde aparecen.

Se eligió relacionar las oraciones con los personajes.

```
{'Acier': ['reminded',
  'Acier',
  'silva',
  'Mereleona',
  'approach',
  'Noelle',
  'offer',
  'word',
  'encouragement',
  'motivate',
  'give',
  'Mereleona',
```

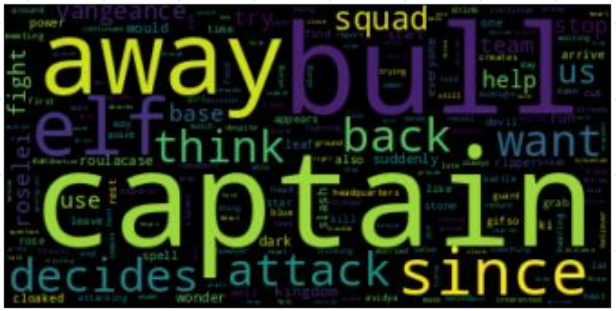
Palabras más mencionadas por Personaje

Asta

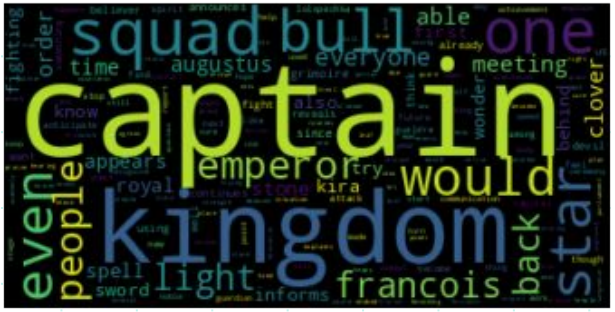


Noelle

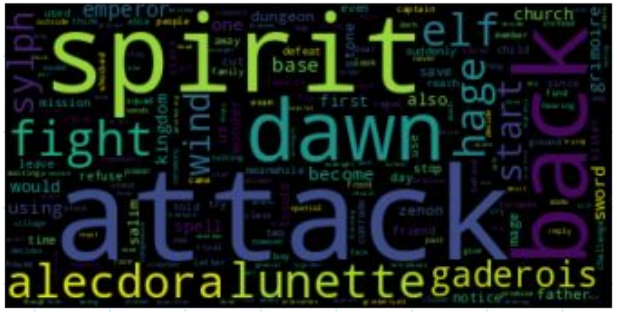
Yami

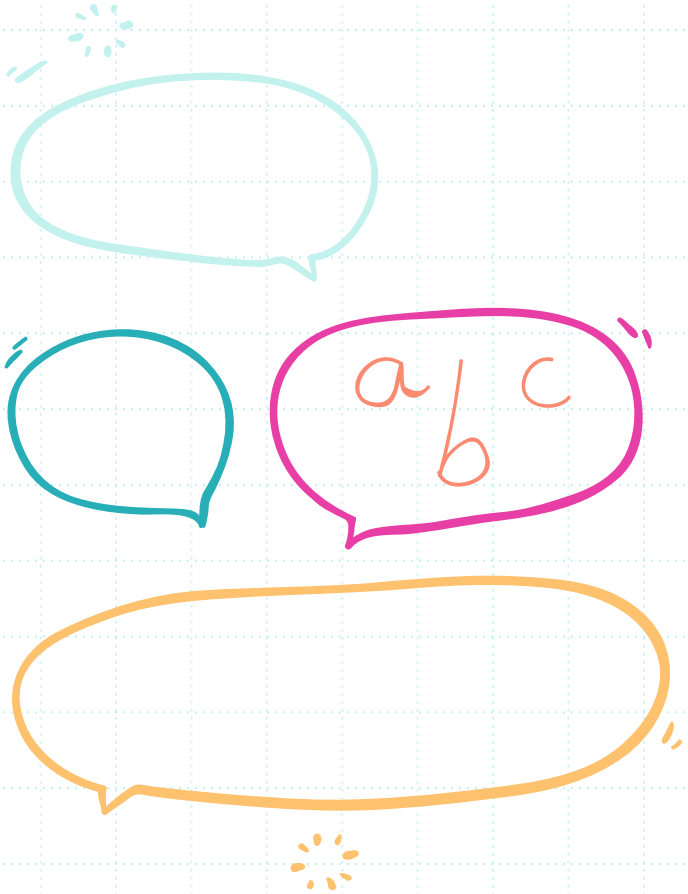


Julius



Yuno





05

Encodeando

Cómo adaptar el corpus para darselo a una red

Encodings

Se aplicó tanto **Hot Enconding** cómo **Count Vectorize**.

Se realizó el Encoding de cada oración, teniendo por cada token 1 si está presente en la oración y 0 si no.

Adicionalmente, se extrajeron los **sujetos de la oración** y se hizo un Encoding de los personajes asociado a cada oración

Asta	Yuno	swing	help	fight
1	1	0	0	1

Asta	Yami	Yuno	Noelle	Julius
1	0	1	0	0

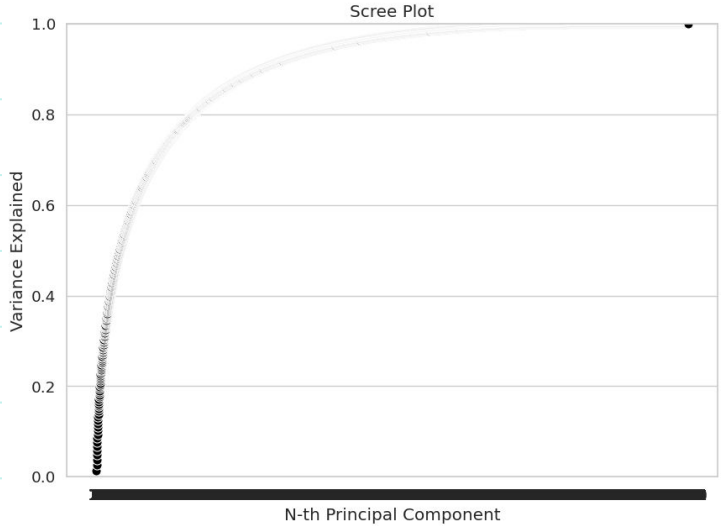
Ajustando Encodings

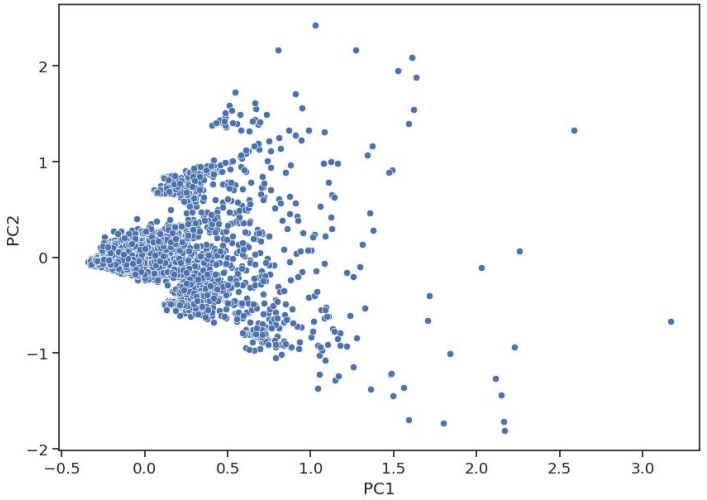
```
[1.0,
 0.5532473212333261,
 0.5431882790290837,
 0.518259348349005,
 0.4277279685108244,
 0.36693636562431664,
 0.23923026459654495,
 0.22086157883227642,
 0.21626940739120928,
 0.2138639842554122,
 0.1758145637437131,
 0.172534441285808,
 0.15613382899628253,
 0.13732779357095998,
 0.13339164662147387,
 0.12989284933304177,
 0.1281434506888257,
 0.1277061010277717,
 0.12355127924775858,
 0.12333260441723158,
 0.10868139077192215,
 0.10736934178876012,
 0.09665427509293681,
 0.09009403017712661,
 0.08768860704132954,
...
 0.0015307238136890444,
 0.0015307238136890444,
```

```
[0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
...
 0.00021867483052700635,
 0.00021867483052700635,
```

- Se encontró una gran diferencia en las medias correspondientes a los encodings.
- Esto corresponde con el gráfico conseguido mediante PCA.
- Para esto se usaron los datos “sucios”, no los procesados que veníamos trabajando.

PCA





Tiene sentido dado los datos “sucios”

Ajustando Encodings

```
[1.0,
 0.5532473212333261,
 0.5431882790290837,
 0.518259348349005,
 0.4277279685108244,
 0.36693636562431664,
 0.23923026459654495,
 0.22086157883227642,
 0.21626940739120928,
 0.2138639842554122,
 0.1758145637437131,
 0.172534441285808,
 0.15613382899628253,
 0.13732779357095998,
 0.13339164662147387,
 0.12989284933304177,
 0.1281434506888257,
 0.1277061010277717,
 0.12355127924775858,
 0.12333260441723158,
 0.10868139077192215,
 0.10736934178876012,
 0.09665427509293681,
 0.09009403017712661,
 0.08768860704132954,
...
 0.0015307238136890444,
 0.0015307238136890444,
```

```
[0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
 0.0,
...
 0.00021867483052700635,
 0.00021867483052700635,
```

- Se decidió eliminar los tokens irrelevantes (cuya media es 0)
- Se decidió remover aquellos tokens cuya media represente un valor mayor a 1.5 veces el desvío estándar promedio

Diferencias entre Hot Encoding y Count Vectorize

Oraciones donde Count Vectorize tiene un valor superior a 1 para Oraciones

```
(array([ 12,   18,   21,   21,   25,   25,   25,   28,   29,   33,   36,
          37,   69,   80,  100,  148,  213,  223,  249,  266,  272,  301,
         304,  317,  339,  367,  373,  373,  410,  424,  430,  438,  473,
         474,  475,  501,  501,  510,  551,  605,  605,  640,  649,  652,
         654,  655,  657,  657,  657,  659,  737,  774,  775,  794,  801,
         804,  819,  876,  886,  896,  901,  906,  906,  906,  906,  906,
         906,  906,  906,  929,  944,  965,  968,  985,  988, 1004, 1043,
        1045, 1089, 1131, 1239, 1249, 1266, 1291, 1291, 1292, 1294, 1325,
        1334, 1351, 1354, 1371, 1371, 1374, 1395, 1396, 1396, 1396, 1406,
        1417, 1432, 1457, 1476, 1510, 1534, 1554, 1572, 1572, 1597, 1601,
        1677, 1691, 1707, 1740, 1773, 1780, 1785, 1801, 1804, 1821, 1822,
        1840, 1840, 1846, 1860, 1872, 1876, 1938, 1943, 1958, 1979, 1983,
        1983, 1983, 1988, 2001, 2002, 2004, 2007, 2007, 2008, 2008, 2010,
        2014, 2014, 2018, 2021, 2024, 2032, 2034, 2035, 2041, 2042, 2042,
        2043, 2045, 2048, 2056, 2056, 2060, 2064, 2067, 2067, 2078, 2086,
        2098, 2105, 2109, 2110, 2112, 2113, 2114, 2121, 2121, 2122, 2128,
        2132, 2136, 2143, 2145, 2146, 2147, 2149, 2154, 2155, 2160, 2170,
        2170, 2178, 2182, 2183, 2195, 2198, 2209, 2210, 2220, 2222, 2227,
        2227, 2229, 2229, 2234, 2240, 2246, 2247, 2253, 2254, 2260, 2281,
        2282, 2282, 2286, 2287, 2289, 2289, 2292, 2292, 2307, 2308, 2308,
        2319, 2323, 2325, 2326, 2326, 2326, 2326, 2327, 2327, 2331, 2335,
        2335, 2335, 2344, 2344, 2344, 2344, 2345, 2358, 2362, 2362, 2362,
        2365, 2366, 2371, 2372, 2378, 2378, 2380, 2382, 2386, 2386, 2388,
        2393, 2401, 2401, 2401, 2414, 2425, 2425, 2426, 2426, 2432, 2434,
        2434, 2434, 2437, 2442, 2444, 2444, 2453, 2454, 2457, 2472, 2473,
...
        1397,  359, 1836,  106, 2374, 4425,  823,  316, 1852, 4919, 3713,
        4091, 4861,  703, 1690, 2881,  702, 3021,  511, 2428, 1745, 3021,
        3508, 3555, 3555, 3555, 2255, 1397, 2428, 4425,  578, 3021, 1745,
        1034, 2513,  702, 4425, 3273,  317, 2513, 3837, 4294, 4899, 1035,
        4605,  347, 1754, 2428,  756, 4960, 1552, 3837, 3065]))
```

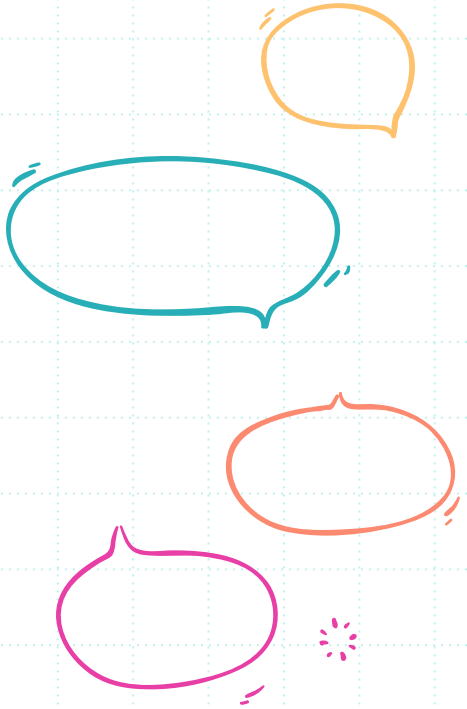
```
(array([  6,   6,   7,  10,  14,  16,  43,  47,  50,  50,  65,
         77,  79,  88, 130, 173, 176, 235, 236, 249, 315, 316,
        353, 354, 369, 381, 394, 395, 397, 423, 460, 544, 574,
        602, 647, 649, 649, 654, 659, 667, 704, 733, 737, 740,
        765, 772, 777, 780, 784, 785, 791, 875, 906, 914, 919,
        930, 1077, 1081, 1126, 1131, 1155, 1167, 1172, 1212, 1249, 1295,
       1309, 1317, 1344, 1366, 1396, 1432, 1457, 1461, 1467, 1476, 1481,
       1503, 1503, 1506, 1528, 1580, 1603, 1607, 1637, 1666, 1703, 1744,
       1802, 1809, 1910, 1916, 1972, 1980, 1981, 1982, 1987, 1994, 2003,
       2005, 2009, 2014, 2035, 2051, 2055, 2056, 2058, 2060, 2061, 2067,
       2086, 2088, 2091, 2108, 2137, 2139, 2142, 2144, 2144, 2156, 2157,
       2161, 2162, 2162, 2170, 2178, 2191, 2195, 2197, 2199, 2201, 2206,
       2217, 2228, 2242, 2256, 2261, 2292, 2296, 2301, 2304, 2305, 2314,
       2323, 2334, 2335, 2338, 2362, 2366, 2386, 2442, 2493, 2496, 2499,
       2504, 2505, 2510, 2513, 2514, 2514, 2524, 2524, 2527, 2528, 2533,
       2536, 2541, 2541, 2542, 2542, 2545, 2554, 2555, 2556, 2563, 2563,
       2565, 2566, 2573, 2574, 2581, 2586, 2597, 2605, 2606, 2613, 2616,
       2620, 2628, 2631, 2649, 2650, 2659, 2663, 2665, 2677, 2717, 2721,
       2722, 2729, 2729, 2730, 2774, 2775, 2780, 2781, 2790, 2823, 2837,
       2842, 2843, 2849, 2855, 2858, 2860, 2861, 2863, 2863, 2864, 2866,
       2866, 2867, 2868, 2874, 2881, 2884, 2893, 2928, 2979, 2984, 2995,
       3007, 3010, 3011, 3012, 3013, 3014, 3014, 3015, 3016, 3017, 3020,
       3056, 3058, 3076, 3140, 3143, 3156, 3156, 3166, 3166, 3167, 3180,
       3182, 3208, 3208, 3229, 3235, 3247, 3261, 3275, 3295, 3304, 3304,
       3309, 3330, 3339, 3340, 3341, 3363, 3372, 3372, 3461, 3489, 3520,
...
         10, 29, 10,  2, 10,  2,  2, 13, 15, 10, 29, 29, 29, 11, 10, 11, 29,
         29, 33,  1, 29, 33, 29, 44, 33,  1,  1, 34, 34,  1,  1, 10, 34, 10,
         22, 22, 44, 19, 26, 26, 44, 25, 26, 29, 44, 29, 18, 26, 24, 14, 18,
         18, 18, 14, 14, 14, 29, 29,  7, 12, 29, 29, 29, 29, 29, 12, 18, 29,
         18, 18, 20]))
```

Oraciones donde Count Vectorize es diferente a Hot Encoding para Personajes

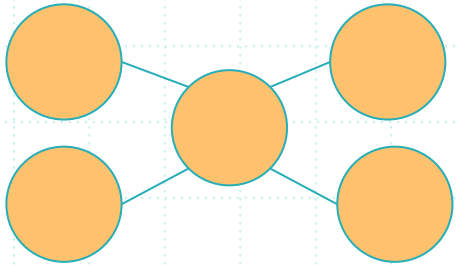
06

Clasificando

Relacionar palabras con personajes



Arquitectura

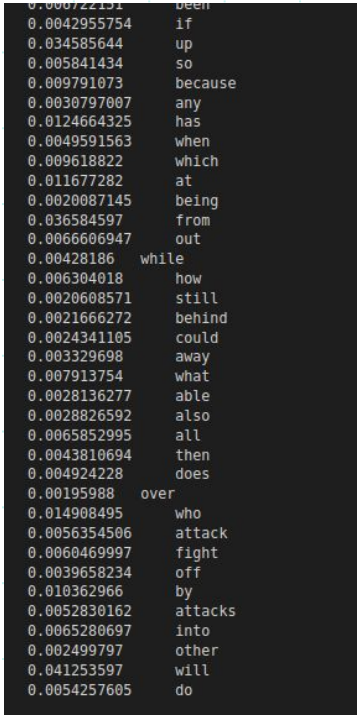


9 capas, función de activación Sigmoidea

Función de activación Softmax en la salida

- La entrada son los personajes como vector
- La salida son los tokens
- Relaciona personajes con tokens
- Las capas intermedias tienen una cantidad de neuronas igual a la cantidad de tokens
- Se probó utilizando RELU y LeakyRELU en las capas intermedias pero sorprendentemente la Sigmoidea dió mejores resultados

Resultados



- Los resultados no fueron buenos
- Se deben limpiar los datos, la mayoría de las probabilidades se las llevan palabras comunes
- Sería interesante hacer este análisis con adjetivos
- Sería útil relacionar palabras entre sí y generar nuevas palabras que sumarizan esa relación

Entrada y Salida Hot Enconded
1000 Épocas
Predicción sobre Yuno