NLP Lab0

Oliver Ramirez

1 Setting up

```
student@docker:/cs/cs159/data/gutenberg$ head chesterton-brown.txt
[The Wisdom of Father Brown by G. K. Chesterton 1914]


I. The Absence of Mr Glass


THE consulting-rooms of Dr Orion Hood, the eminent criminologist
and specialist in certain moral disorders, lay along the sea-front
at Scarborough, in a series of very large and well-lighted french windows,
which showed the North Sea like one endless outer wall of blue-green marble.
```

1.1 SpaCy Experiment

```
import discord
>>> from spacy.lang.en import English
>>> nlp_model = English()
>>> test_str = """
...
... Hey oliverRamirez4!
...
... A third-party GitHub Application (Docker Inc) with the following permissions:
...
...    - View your email addresses
...
... Was recently authorized to access your account.
...
... Visit https://nam04.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgithub.com%2
...
... To see this and other security events for your account, visit https://nam04.safelink
...
... If you run into problems, please contact support by visiting https://nam04.safelinks
...
```

```
... Thanks,
... The GitHub Team
... """
>>> spacy_output = nlp_model(test_str)
>>> for piece in spacy_output:
...     if piece.like_url:
...             print(piece)
...
https://nam04.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgithub.com%2Fsettings%2
https://nam04.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgithub.com%2Fsettings%2
https://nam04.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgithub.com%2Fcontact&d
>>>
```

1.1.1 reflection

The code matched URLs perfectly. It sorted out the 3 URLs and grabbed the
correct text that contained all relevant URLs.

2 Notes

In reflection of the regEx activity, I thought that all of the problems went very
smoothly except for the ABBA problem. I had difficulty figuring out that I
need to use the start of line operatorˆto filter out all of the abba patterns. I am
proud of using the look ahead operator and the backreference problem as well.