Location Project

Ghalia Alshanbari, Jena Jordahl, Bhaskar Abbireddy

Problem Statement

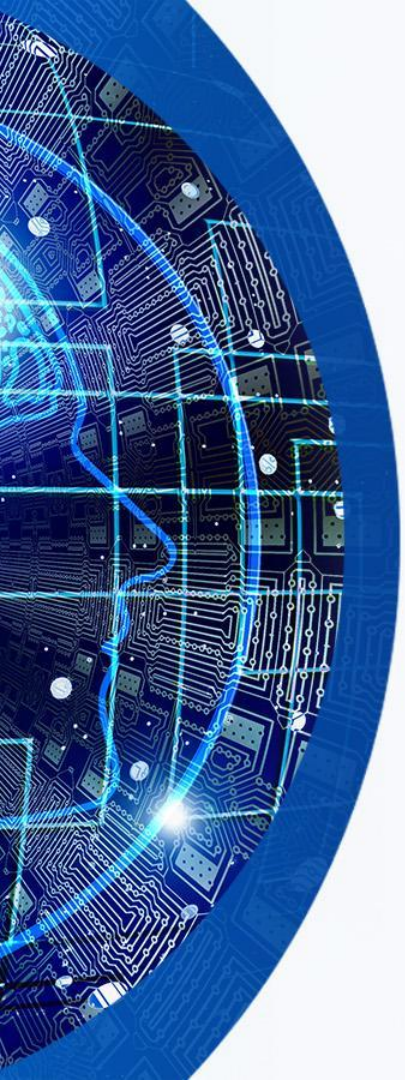
Background: The location of a story is not always explicitly given in the text.

Hypothesis: New transformer NN might be able to tell us where a story takes place.

Goal: Use text classification and graph theory to determine the location where a scene takes place in a fiction or nonfiction story

Our method preprocessed stories into chapters, paragraphs, and sentences. Then grouped those sentences into story scenes based on similarity. Finally, we predicted the location of the scene by inference. We used Bert to predict locations.

As part of our inference process, we generated summary text based on location specific sources. We trained an LSTM as well as a transformer to generate location summary text.



Approach

- Train NNs with game of thrones script and Gettysburg Battle Texts from gutenberg press
- Leveraged SCC, Colab and Local machines to run our code as we hit multiple processing issues
- Visualized the significance of characters in a script/scene via word clouds
- Graphed using networkx and pyvis a forced directed graph to break related texts into scenes

Processed the script from “Gettysburg”

The movie is based on the book Killer Angels

https://civilwarcycling.com/index/gettysburg-maps-hal-jespersen/

Focused on Processing the “Game of Thrones” Script

- The Kingdom of the North, ruled by House Stark of Winterfell
- The Kingdom of the Isles and the Rivers, ruled by House Hoare of Harrenhal
- The Kingdom of Mountain and Vale, ruled by House Arryn of the Eyrie
- The Kingdom of the Rock, ruled by House Lannister of Casterly Rock
- The Storm Kingdom, ruled by House Durrandon of Storm's End
- The Kingdom of the Reach, ruled by House Gardener of Highgarden
- The Principality of Dorne, ruled by House Martell of Sunspear.[S 1]

From Wikipedia: Winterfell

We Selected 8 Places:

Westeros

- Winterfell
- Kings Landing
- Harrenhal
- Iron Islands

Essos

- Valyria
- Pentos


White Woods
Lands of Always Winter
Giant's Lake
Shivering Sea
Essos
Westeros
Sunset Sea
Near Mossovy
Far Mossovy
Dothraki Sea
Grey Waste
Gulf of the Winged
Sunrise Sea
Yi Ti
Valyria
Summer Sea
Shadow Lands
Jade Sea
Ulthos
The Southern Wall
The Cursed Lands
Summer Isles
Karakanda
Sothoryos
Monkey Island
Mermaid Sea
Cannibal Islands
Black March
Ulthosi Ocean
Novii Empire
Lizard Isles
Kissangari
Serpent Sea

Obstacles

SCC Access to GPUs

SCC would not recognize the GPU

Likely there is a tutorial for GPU use

Requesting multiple GPUs requires separating the model to each one

https://huggingface.co/docs/transformers/perf_train_gpu_one

,CUDA_LAUNCH_BLOCKING=1 example

```
There are 2 GPU(s) available.
We will use the GPU: Tesla V100-SXM2-16GB
```

Bugs in COLAB, PyCharm

'_AxesStack' object is not callable while using networkx to plot

Could not plot the entire link because one name is not a string or an int .In 23,910 lines, did not find the one…

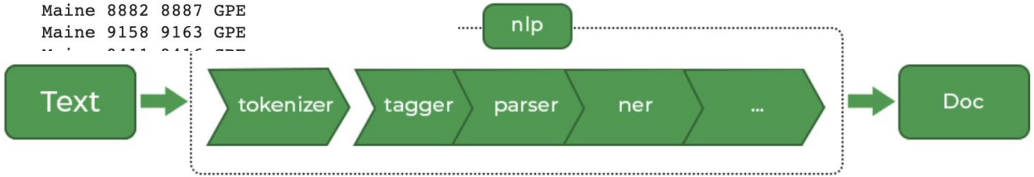
```
-> 233     assert isinstance(n_id, str) or isinstance(n_id, int)
   234     if label:
   235         node_label = label
```

AssertionError:

```
RuntimeError: CUDA error: CUDA-capable device(s) is/are busy or unavailable
CUDA kernel errors might be asynchronously reported at some other API call,so the stacktrace
below might be incorrect.
For debugging consider passing CUDA_LAUNCH_BLOCKING=1.
```

Parser by Spacey

```
Robert E 162 170 PERSON
Lee 173 176 PERSON
Virginia 771 779 GPE
Maryland 869 877 GPE
Pennsylvania 887 899 GPE
Lee 938 941 PERSON
Abraham Lincoln 1131 1146 PERSON
Lee 1219 1222 PERSON
Washington 1315 1325 GPE
Lee 1583 1586 PERSON
Harrison 2184 2192 PERSON
Harrison 2297 2305 PERSON
Jeb Stuart 4062 4072 PERSON
mule 4509 4513 PERSON
Jeb Stuart's 4711 4723 PERSON
Sorrel 5054 5060 GPE
Lee 5128 5131 PERSON
Eleventh 5715 5723 PERSON
Harrison 5952 5960 PERSON
Mississippi 6042 6053 GPE
George Meade's 6425 6439 PERSON
Harrison 6461 6469 PERSON
George Meade 6529 6541 PERSON
Pennsylvania 6543 6555 GPE
Chamberlain 7359 7370 PERSON
bucko 7433 7438 GPE
Maine 8882 8887 GPE
Maine 9158 9163 GPE
```

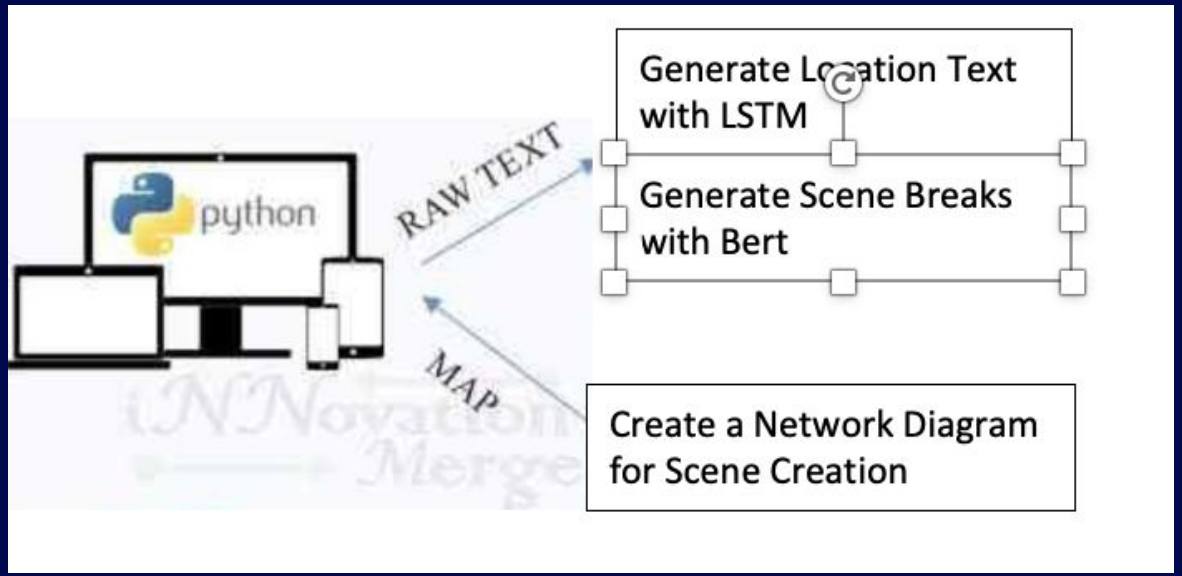


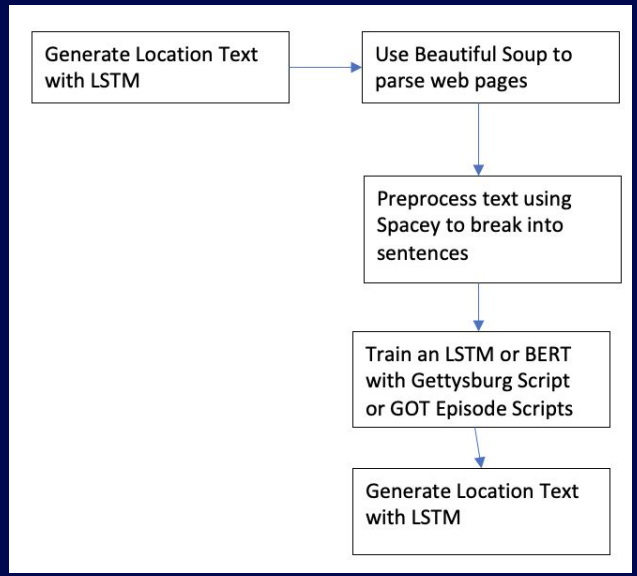
https://www.geeksforgeeks.org/tokenization-using-spacy-library/

Intermediate steps for tokenization

Methodology

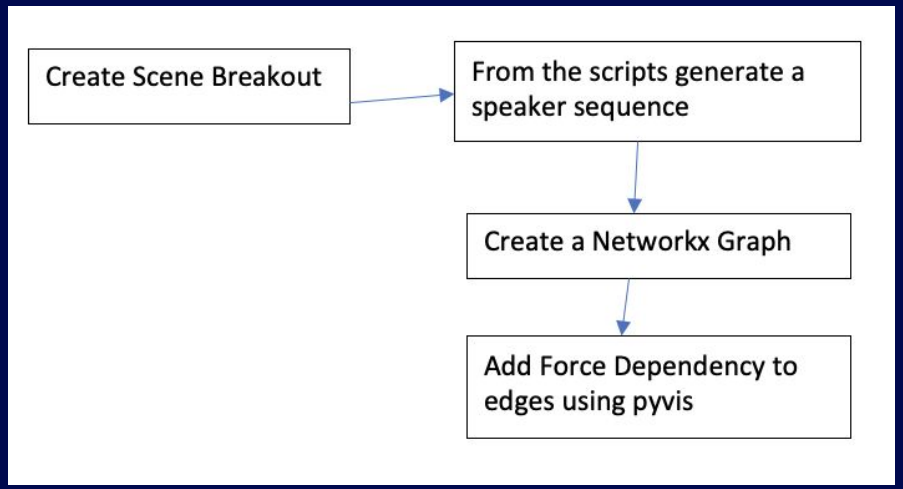
Directly Generating Text & Scene Breakout

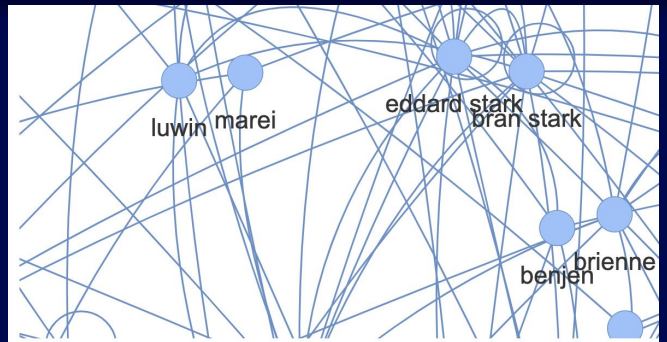


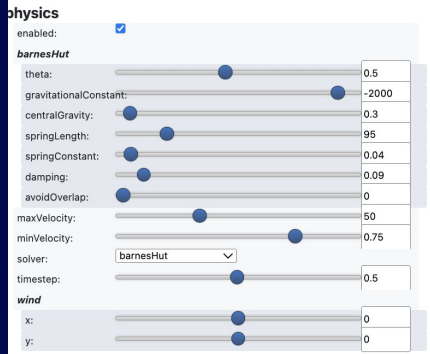


Finding Scene Breaks

Networkx & PyVis







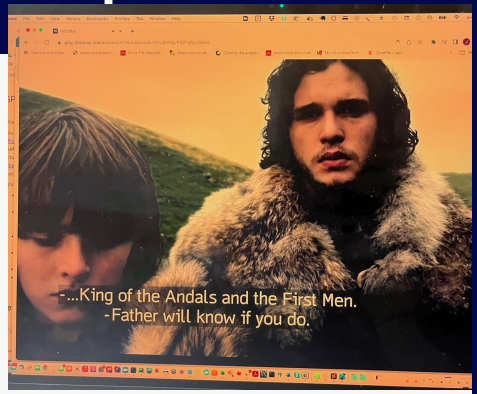
Bert Scene Breakout

Captures Different Conversations

Actual Scenes

List of Scenes BERT pretrained

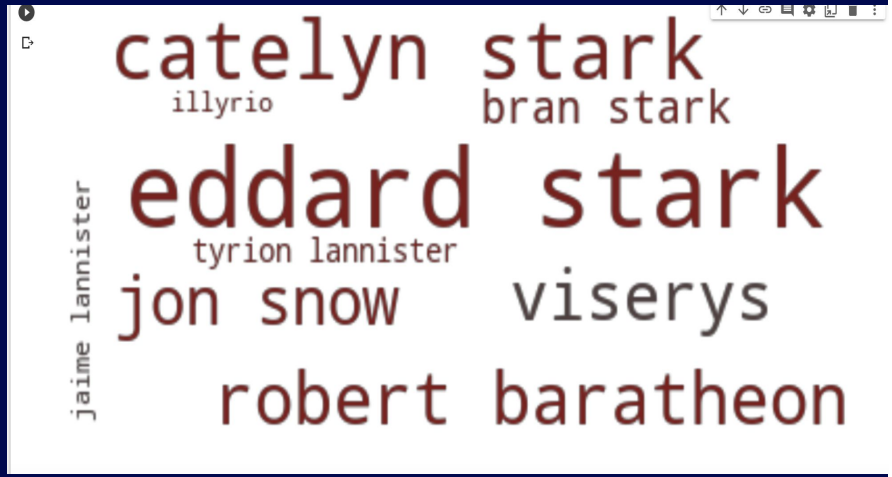
```
    result = torch.argmax(outputs.logits)

    if result.item() == 0:
      inter_result.append(sentences[i+1])
    else:
      result_set.append(inter_result)
      print(i,inter_result)
      inter_result = []
  else:
    result_set.append(inter_result)

print(result_set)
```



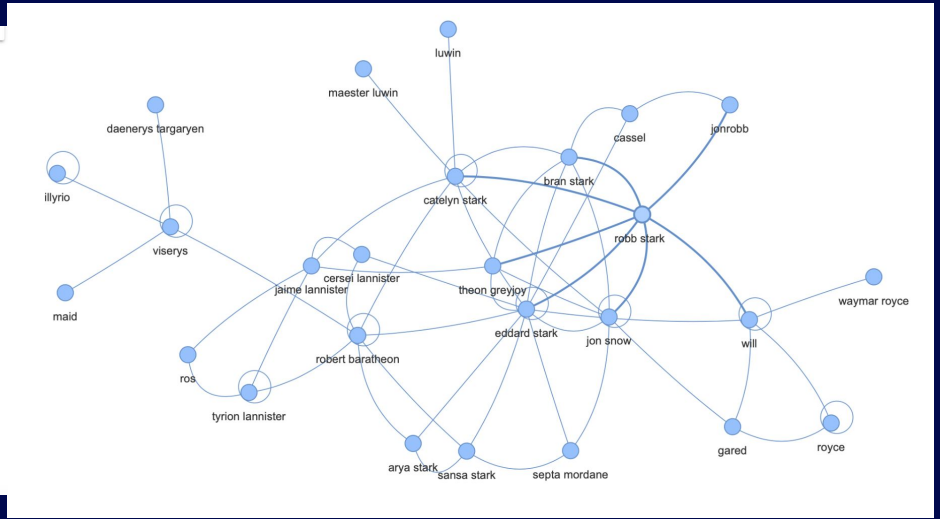

```
3 ['What d'you expect? They're savages One lot steals a goat from another lot and before y
36 ['We should head back to the wall', 'Do the dead frighten you?', 'Our orders were to tr
37 ['Don't look away']
38 ['King of the Andals and the First Men']
46 ['Father will know if you do', 'Lord of the Seven Kingdoms and protector of the realm,
48 ['Is it true he saw the White Walkers?', 'The White Walkers have been gone for thousand
49 ['So he was lying?']
56 ['A madman sees what he sees', 'What is it?', 'Mountain lion?', 'There are no mountain
59 ['There are no direwolves south of the Wall', 'Now there are five', 'You want to hold i
72 ['Where will they go? Their mother's dead', 'They don't belong down here', 'Better a qu
```

Results

Word cloud



Graph



Models for Gen Text

Created and Saved LSTM

```
vocabulary sizes: 5904,5904
token2int['_PAD'] 5903
token2int['_UNK'] 5902
```

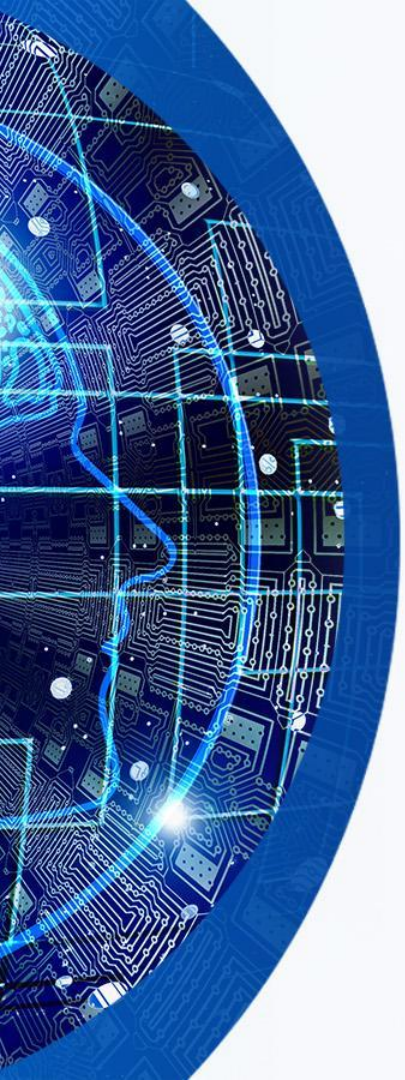
```
[ ]
```

```
vocabulary sizes: 3909,3909
token2int['_PAD'] 3908
token2int['_UNK'] 3907
```

Script Appropriate

<s> i 'm sorry i 'm not going back to king in the south ! ” it 's the only way to die . " i 'll have a baby , you 're the only person left in the city , and we have a choice , i 'll be the first of my own father , the queen of the iron islands . ” and i 'll be sure . . </s>

kings landing is east of the dead and we 're not going home , but we 'll be the first man in westeros ? ” it is the only one i ever thought to confess , but you 'll have a choice , you know . ! ” it is the best who know what i 've seen the people of westeros and their lives are defeated and _UNK . </s>



Results of Location Classification

Prompt Completion ACCURACY, determined subjectively:

Gettysburg: 20 correct; 52 incorrect; correct percentage 27%
Gettysburg: 13 correct; 59 incorrect; correct percentage 18%
Conclusion: LSTMS are not good at predicting location.

What we would add

Trying other model types

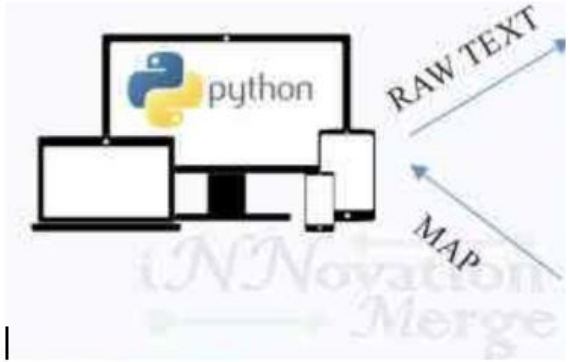
The extra credit will be if we can predict location relationships from the surrounding text to pinpoint where the scene is taking place.

For example: Show a map that represents the scene

Future Methodology

NN Generated Text

NN Classification



Generate Location Text with LSTM

Generate Location Text with Bert

Bert NLI classification

Create a Map

Resources/References

- https://networkx.org/documentation/stable/reference/generated/networkx.convert_matrix.from_pandas_edgelist.html
- https://towardsdatascience.com/visualizing-networks-in-python-d70f4cbeb259
- https://loremaps.azurewebsites.net/Maps/GoT
- https://medium.com/@pacificlion/visualize-network-data-9df1d4c796a3
- https://link.springer.com/content/pdf/10.1007/s13278-021-00777-5.pdf?pdf=button
- https://medium.com/spatial-data-science/how-to-extract-locations-from-text-with-natural-language-processing-9b77035b3ea4
- https://scholar.smu.edu/cgi/viewcontent.cgi?article=1214&context=datasciencereview
- https://towardsdatascience.com/easiest-way-to-plot-on-a-world-map-with-pandas-and-geopandas-325f6024949f https://www.geeksforgeeks.org/extracting-locations-from-text-using-python/
- https://www.youtube.com/watch?v=awB-isTpWrU
- https://www.gq-magazine.co.uk/article/game-of-thrones-map
- https://www.etsy.com/fi-en/listing/270976211/1863-map-of-gettysburg-civil-war?ga_order=most_relevant&ga_search_type=all&ga_view_type=gallery&ga_search_query=gettysburg+map&ref=sr_gallery-1-13&sts=1&organic_search_click=1
- https://civilwarcycling.com/index/gettysburg-maps-hal-jespersen/