Choose Your Own Adventure Story

DSCI6004: Jonathan Hilgart
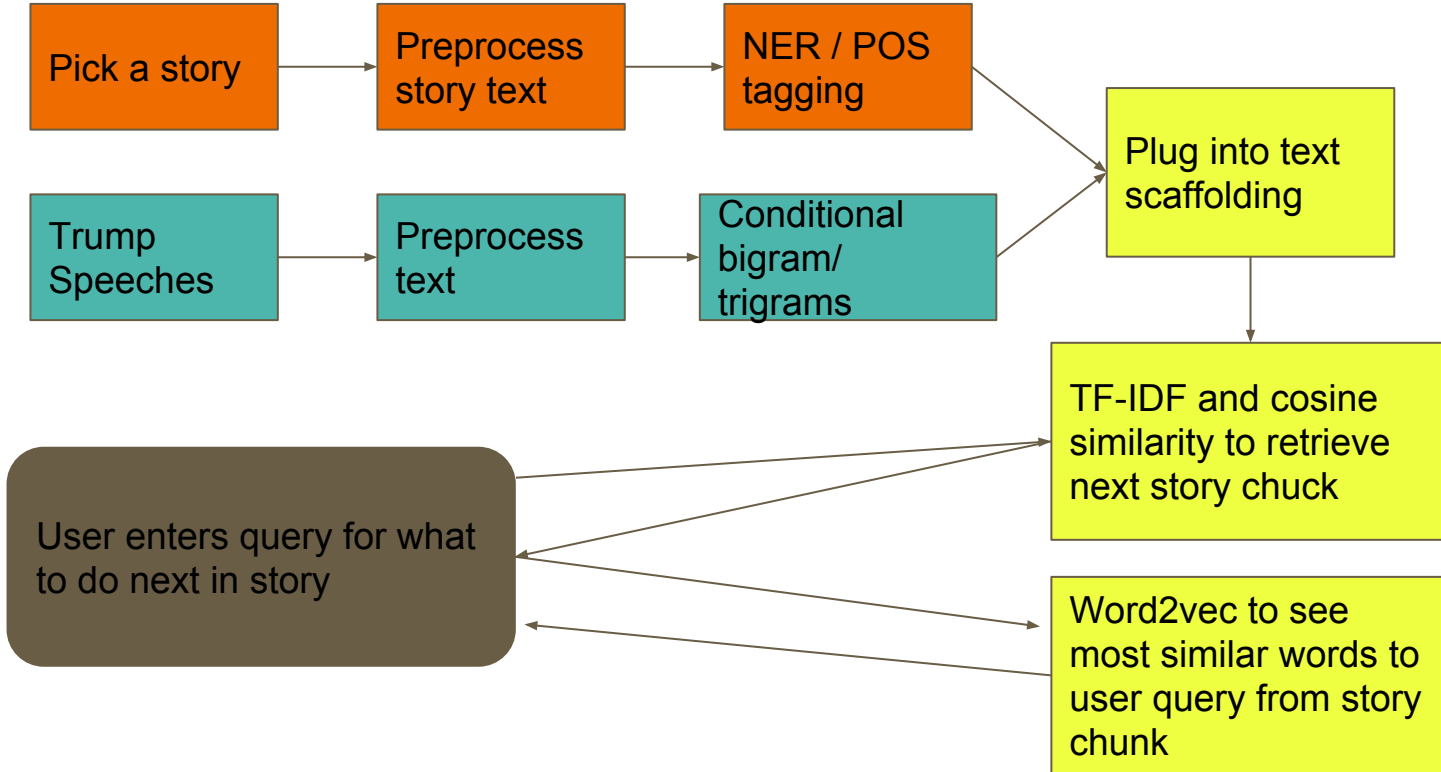
Overview

- Mad-libs style adventure story
  - Alice in Wonderland
  - Ulysses
  - Grimms Fairy Tales
- User picks three items to use throughout the story
- Can you survive?

https://blog.codinghorror.com/choosing-your-own-adventure/

Architecture

Pick a story → Preprocess story text → NER / POS tagging

Trump Speeches → Preprocess text → Conditional bigram/trigrams

Plug into text scaffolding

TF-IDF and cosine similarity to retrieve next story chuck

User enters query for what to do next in story

Word2vec to see most similar words to user query from story chunk

Story Chunks

- Goal: Pick story chunk most relevant to user query for the next part of the story

```python
return ['You have the feeling that something bad was going to happen so you pull out your pocketknife. {} looks at you and asks,"So what do you\
think you are going to do with that? That is not going to save you. You notice {} in the distance and you yell to trump "Run for it!" . Where sho
.format(protag,pl_one),

'All of a sudden, you remember that you have not eaten lunch. "Silly me!" you think to yourself as your pull out your half-eaten-sandwich.\
You chew happily for a while before deciding it was time to move on. "Time to go you mention to Trump". As you walk you suddenly come across {} \
who asks "What are ye doing here?" What do you say?'.format(antag),

' After talking with {} you asked "What do we do around here?". {} replies, "Currently, I am avoiding {} because they are looking\
to round up everyone who speaks {}!" "Oh no!" You think to yourself "They are going to capture Trump!"Trump turns and says {} \
"Wow, he really has a way with words," you think. Would should you do next?'.format(protag,protag,antag,lang_one,trump_speech),   # protag, protag

' It starts to get dark out so you pull out your flashlight. "Good thing I brought this", you think. You glance around and see some {}.\
What are these things you ask Trump? He replies {} . "Again, your not making sense!" you state as your start to get angry. \
Regardless, you continue forward towards the {} until you reach it. What should you do next?'.format(pn_one,trump_speech,pn_one),
#plural common noun, trump , same plural noun
' "Why did I even bring this with me?" You wonder looking at the broken-lightblub. You continue staring for the next 30 minutes, wondering \
if there was a deeper meaning to the object. You decide you had enough and you {} the lightblub. Great, now what?'.format(vb_one),

' "It is time to explore!" you announce to {} . With that, you jump on your bike, put Trump on the handlebars, and start going. \
"Wait!" yells {} "You are going to need this," she says handing you a {}. What do you do next? '.format(protag,protag,sn_one),

' "It is time to explore!" you announce to {} . With that, you jump on your skateboard, before realizing that Trump was not going to fit on your
"Hmmm," you think. Maybe I better try something else. What do you do next?'.format(protag,sn_one)
]
```
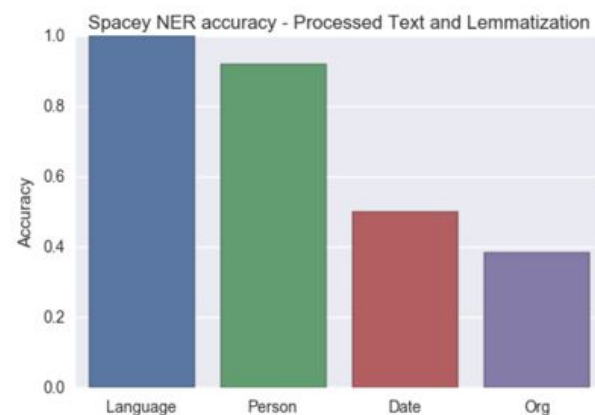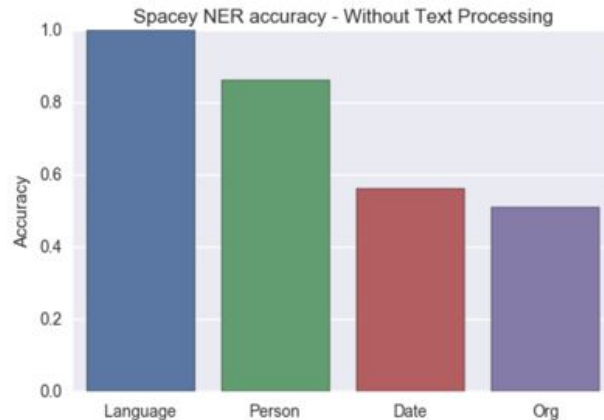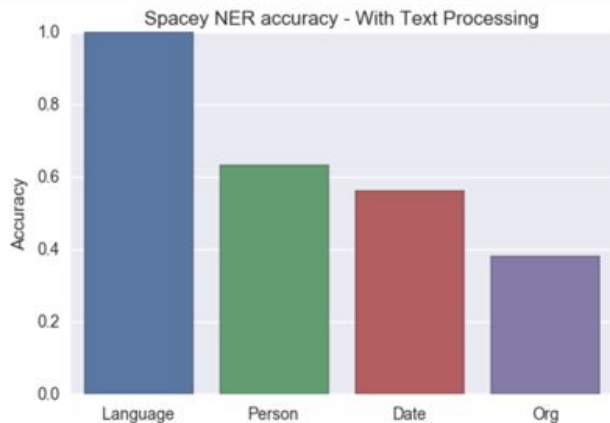
Techniques used

- Use NER, POS to pull out characters, places, verbs ..etc from the story
  - Plug into text scaffolding

- Probabilistic language generation for your companion, Trump
  - Unigram/Bigram/Trigram model

'Feeling upset, you take out your half-eaten-sandwich from lunch and throw it on the ground. You watch the jelly ooze onto the ground with disappointment.   Trump turns to your with bulging eyes and a hand over his stomach and says He stopped . And what are we going to do this thing is if you want to fix our rigged system so we have . Record , record , putting the american people can't do that one ? . Now what? '

- Information retrieval model for the next part of the story
  - Tf-idf with cosine similarity (also tested Jaccard Similarity).
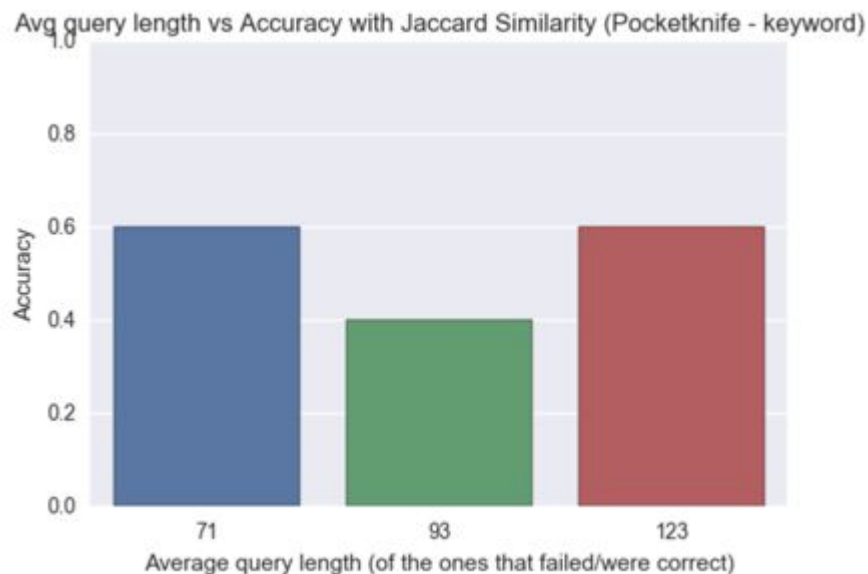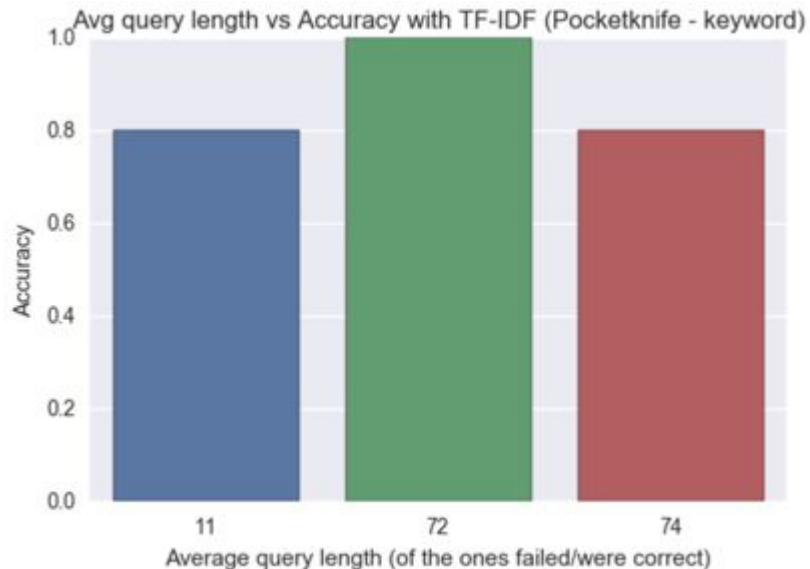 - Word2vec to see the related words in the story chunk section

Evaluation of NER

- Text processing consists of removing special characters and all punctuation except for periods, exclamation marks, and questions.
- Some entities may not make sense in categories with low accuracy. Therefore, used most common entity in each category to mitigate this chance
- Used lemmatization due to the increase in the person's category accuracy.

Evaluation of Information Retrieval system

- Evaluate to see if story chunk containing keyword is retrieved.
- Test TF-IDF with cosine similarity or Jaccard Similarity.

Avg query length vs Accuracy with TF-IDF (Pocketknife - keyword)

Avg query length vs Accuracy with Jaccard Similarity (Pocketknife - keyword)

Live demo!