

SONNET SPOOFER  
by Shari Kuroyama and Carly Robison

---

## 1 Tokenizing

### Methods Used

What methods did you use and try to tokenize the sonnets?

- First we tried the naïve tokenization of using each word as a token. With enough states and iterations, this actually learned fairly well.

### Modifications

Did you have to make changes to the way you tokenized after running the algorithm and seeing the results?

- We thought about splitting each word into syllables in order to keep to the 10-syllable format of a sonnet. However, we got around that by counting the number of syllables generated per line, and generating new words if the last word was too long.
- We also thought about trying to train using bigrams of words as tokens; but again, the naïve implementation worked fairly well.

## 2 Algorithm

### Packages Used

What packages did you use for the algorithm?

### Parameters Used

What decisions did you have to make when running the algorithm and what did you try? e.g number of states

### Effect On Sonnets

How did this affect the sonnets that were generated

## 3 Poetry Generation

### How?

How did you generate your poem?

### How to get it to look like a sonnet?

How did you get your poem to look as much like a sonnet as possible?

### Do they make sense?

What makes sense/what doesn't about the sonnets generated?

## 4 Visualization and Interpretation

### Hidden State Meaning

For at least 5 hidden states give a list of the top 10 words that associate with this hidden state and state any common features these groups.

### Properties of Hidden States

What are some properties of the different hidden states? e.g. Correlation between hidden states and syllable counts, connotations of words, etc.

### Visual representation

Make a visual representation of the correlation between states and words

## 5 Additional Goals

e.g. rhyme/meter

## 6 Extra Credit: RNN

What did you use to collect more data and what packages did you use for RNN/LSTM? Compare/contrast the effect of these algorithms to HMM and why you think they were better or worse

## 7 Conclusion

### Division of work

### Discoveries

What are your conclusions/observations about the models you used and the sonnets generated?

### Challenges

### Concluding Remarks

- hey there