Introduction to Computation for the Humanities and Social Sciences

CS 3
Chris Tanner
“Wu-Tang Clan Ain’t Nuthin to [Compare] With”
— Wu-Tang Clan
Lecture 13

- Sets
- Nested Iteration
- Text Analysis: readability
  - contributing factors
  - measures
New Data Structure!

- very similar to Lists

- A Set is an unordered collection of items (can store any data type)

- Collections which don’t preserve the order in which elements are added are called unordered

- Doesn’t support indexing

```python
a = {1,2,3}
a.add(4)
if 2 in a:
    print(a)
```
Sets

New Data Structure!

- **Sets** have **no indices** and no order!

```python
students = set()
students.add("jackie")
students.add("emily")
students.add("hank")
students.add("emily")
students[0] # CRASHES
```

- So, you can’t access specific items by its index. If you need to maintain an order for your items, use a **List**.
New Data Structure!

- very similar to Lists, but
- is incapable of storing duplicate items (maintains just 1 copy)

```
students = set()
students.add("jackie")
students.add("emily")
students.add("hank")
students.add("emily")
print(students)
```

```python
>>> students = {'emily', 'jackie', 'hank'}
```
Sets

New Data Structure!

• checking if an item is contained in the Set happens instantly (doesn’t need to check each item one by one like Lists)

```python
students = set()
students.add("jackie")
students.add("emily")
students.add("hank")
students.add("emily")

if cur_student in students:
    print(student + " is present!")
else:
    print(student + " is NOT present!")
```
When to use a `Set()`?

• Use a `Set()` when you only care about storing a collection of unique values (i.e., don’t care about duplicates) and the order doesn’t matter to you.

• Pros: super simple, fast data structure; never need to worry about indices.

• e.g., store license plate #s or list of senator names.
Lecture 13

- Sets
- Nested Iteration
- Text Analysis: readability
  - contributing factors
  - measures
Nested Iteration

Iteration

• Remember, iteration allows us to loop through a given section of code (e.g., a specified number of times or through each item in a particular data structure that is Iterable)

• Just like how we can nest multiple if-statements, we can nest multiple iterations!
Nested Iteration

Example

cs3_students = ["chris", "riya", "laurel", ...]
eng16_students = ["piyush", "anqi", "jana", ...]

for cs3_student in cs3_students:
    for eng16_student in eng16_students:
        if cs3_student == eng16_student:
            print("classes have a student with same name!")
Lecture 13

- Sets
- Nested Iteration
- Text Analysis: readability
  - contributing factors
  - measures
Problem: You want to promote reading for fun with your students and want to point them to grade-level appropriate books without having to read every single one.
Problem: Your company is launching a new product and you are tasked with determining whether the instruction manual will be readable by its customers.
Problem: You are writing a speech for a political candidate that will reach an audience of millions. How do you ensure the message is understood by as many people as possible?
It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair.

Charles Dickens - A Tale of Two Cities
Lecture 13

- Sets
- Nested Iteration
- Text Analysis: readability
  - contributing factors
  - measures
What are some possible factors that influence the difficulty/complexity of a given piece of text?
Text Readability: Factors

Vocabulary Size

![Graph showing vocabulary size over age for different reading ability groups.](image-url)
Text Readability: Factors

Vocabulary Size

# OF UNIQUE WORDS USED WITHIN ARTIST’S FIRST 35,000 LYRICS

Notes/sources:
(1)(2) I used the first 5,000 words for 7 of Shakespeare’s works: Hamlet, Romeo and Juliet, Othello, Macbeth, As You Like It, Winter’s Tale, and Troilus and Cressida. For Melville, I used the first 35,000 words of Moby Dick.
Text Readability: Factors

Vocabulary Size

# OF UNIQUE WORDS USED WITHIN ARTIST’S FIRST 35,000 LYRICS

Notes/sources:
(1)(2) I used the first 5,000 words for 7 of Shakespeare’s works: Hamlet, Romeo and Juliet, Othello, Macbeth, As You Like It, Winter’s Tale, and Troilus and Cressida. For Melville, I used the first 35,000 words of Moby Dick.
Shakespeare loses to the Wu-Tang Clan in vocabulary duel

https://www.pbs.org/newshour/arts/data-scientist-pits-shakespeare-wu-tang-clan-battle-words
Her antiquity in preceding and surviving succeeding tellurian generations: her nocturnal predominance: her satellitic dependence: her luminary reflection: her constancy under all her phases, rising and setting by her appointed times, waxing and waning: the forced invariability of her aspect: her indeterminate response to inaffirmative interrogation: her potency over effluent and reflu ent waters: her power to enamour, to mortify, to invest with beauty, to render insane, to incite to and aid delinquency: the tranquil inscrutability of her visage: the terribility of her isolated dominant resplendent propinquity: her omens of tempest and of calm: the stimulation of her light, her motion and her presence: the admonition of her craters, her arid seas, her silence: her splendour, when visible: her attraction, when invisible.

James Joyce - Ulysses
God of heaven there's nothing like nature the wild mountains then the sea and the waves rushing then the beautiful country with the fields of oats and wheat and all kinds of things and all the fine cattle going about that would do your heart good to see rivers and lakes and flowers all sorts of shapes and smells and colours springing up even out of the ditches primroses and violets nature it is as for them saying there's no God I wouldn't give a snap of my two fingers for all their learning why don't they go and create something I often asked him atheists or whatever they call themselves go and wash the cobbles off themselves first then they go howling for the priest and they dying and why why because they're afraid of hell on account of their bad conscience ah yes I know them well who was the first person in the universe before there was anybody that made it all who ah that they don't know neither do I so there you are they might as well try to stop the sun from rising tomorrow the sun shines for you he said the day we were lying among the rhododendrons on Howth head in the grey tweed suit and his straw hat the day I got him to propose to me yes first I gave him the bit of seedcake out of my mouth and it was leapyear like now yes 16 years ago my God after that long kiss I near lost my breath yes he said I was a flower of the mountain yes so we are flowers all a woman's body yes that was one true thing he said in his life and the sun shines for you today yes that was why I liked him because I saw he understood or felt what a woman is and I knew I could always get round him and I gave him all the pleasure I could leading him on till he asked me to say yes and I wouldn't answer first only looked out over the sea and the sky I was thinking of so many things he didn't know of Mulvey and Mr Stanhope and Hester and father and old captain Groves and the sailors playing all birds fly and I say stoop and washing up dishes they called it on the pier and the sentry in front of the governors house with the thing round his white helmet poor devil half roasted and the Spanish girls laughing in their shawls and their tall combs and the auctions in the morning the Greeks and the Jews and the Arabs and the devil knows who from all the ends of Europe and Duke street and the fowl market all clucking outside Larby Sharons and the poor donkeys slipping half asleep and the vague fellows in the cloaks asleep in the shade on the steps and the big wheels of the carts of the bulls and the old castle thousands of years old yes and those handsome Moors all in white and turbans like kings asking you to sit down in their little bit of a shop and Ronda with the old windows of the posadas 2 glancing eyes a lattice hid for her lover to kiss the iron and the wineshops half open at night and the castanets and the night we missed the boat at Algeciras the watchman going about serene with his lamp and O that awful deepdown torrent O and the sea the sea crimson sometimes like fire and the glorious sunsets and the figtrees in the Alameda gardens yes and all the queer little streets and the pink and blue and yellow houses and the rosegardens and the jessamine and geraniums and cactuses and Gibraltar as a girl where I was a Flower of the mountain yes when I put the rose in my hair like the Andalusian girls used or shall I wear a red yes and how he kissed me under the Moorish wall and I thought well as well him as another and then I asked him with my eyes to ask again yes and then he asked me would I yes to say yes my mountain flower and first I put my arms around him yes and drew him down to me so he could feel my breasts all perfume yes and his heart was going like mad and yes I said yes I will Yes
Text Readability: Factors

Text Length

GeProvidence Retweeted

Rhode Island History @RIHistory · Oct 16

The RIHS is loaning our original 1643 copy of Roger Williams’s “A Key Into the Language of America” to the wonderful @rowi1636 for their event on 10/20, & you’re invited to come check it out!

Text Readability: Factors

Subject Matter

The Constitution of the United States of America (1787) – Full Text

Preamble

We the People of the United States, in Order to form a more perfect Union, establish justice, insure domestic Tranquility, provide for the common defense, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity, do ordain and establish this Constitution for the United States of America.

Article I

Click to view free lesson — Congress and the Constitution

Section 1. All legislative powers herein granted shall be vested in a Congress of the United States, which shall consist of a Senate and House of Representatives.
Lecture 13

- Sets
- Nested Iteration
- Text Analysis: readability
  - contributing factors
  - measures
Text Readability

Flesch-Kincaid

\[ 0.39 \left( \frac{\text{total words}}{\text{total sentences}} \right) + 11.8 \left( \frac{\text{total syllables}}{\text{total words}} \right) - 15.59 \]

- Output is a grade level
- Correlates very well with actual grade level comprehension
- Used very commonly
- Requires computing the number of syllables, which is hard to do very precisely
Text Readability

Dale-Chall

\[ 0.1579 \left( \frac{\text{difficult words}}{\text{words}} \times 100 \right) + 0.0496 \left( \frac{\text{words}}{\text{sentences}} \right) \]

- Includes a list of 3000 words that 4th graders could reliably understand

- Estimates the number of difficult words, by checking if words do not belong to the list of ‘easy words’

- Even more reliable than Flesch-Kincaid, commonly used with modern scientific research
Text Readability

Coleman-Liau

CLI = 0.0588L - 0.296S - 15.8

L: Average number of letters per 100 words
S: Average number of sentences per 100 words

• Motivation: Counting syllables is difficult to generalize
• Breaks text into 100-word chunks
• Computes averages over these word-chunks
LAB TIME