Computing Readability of Texts with Iteration
Interested Topics Survey

• We’ll choose topics to teach and build activities around the responses

• Please fill it out!

• https://goo.gl/forms/FZ9OpQDxm2hdNeTR2
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.
Text Readability

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
Factors that influence Readability
Factors that influence Readability

Vocabulary Size
Factors that influence Readability

**Vocabulary Difficulty**

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 refluent 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.
God of heaven theres 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 theres no God I wouldnt give a snap of my two fingers for all their learning why dont 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 theyre 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 dont 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 womans 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 wouldnt answer first only looked out over the sea and the sky I was thinking of so many things he didnt 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 else 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 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
Factors that influence Readability

Text Length
Factors that influence Readability

Subject Matter
Readability Measures
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
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 belong to the list of ‘easy words’

- Even more reliable than Flesch-Kincaid, commonly used with modern scientific research
Coleman-Liau

\[ CL = 0.0588L - 0.296S - 15.8 \]

- **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
Iteration with List Expressions
Iteration

• Often, we want to repeat a similar computation step a number of times on slightly different data

• To create an iterative process, we can make use of loops

• There are two types of loops we’ll talk about in class, **for loops** and **while loops**

  • **For loops**: Perform the computation step *for* a specific number of times

  • **While loops**: Perform the computation step *while* a given condition is True

• For today we’ll focus on a specific type of for-loop, called list expressions
Iteration

Input Data

Loop

Output: collection/summary
Data types that you can iterate over are called *iterables*. List, Sets, Dictionaries, Tuples are iterables, but so are strings, input files and many other types.
List Expressions

- List expressions require an input iterable and output a list of the same length as the iterable.

- Each item in the output list is a result of a computation over its corresponding item in the input iterable.

```plaintext
input iterable:  
<table>
<thead>
<tr>
<th>item₁</th>
<th>item₂</th>
<th>item₃</th>
<th>item₄</th>
<th>item₅</th>
</tr>
</thead>
</table>

output list:  
| fn(item₁) | fn(item₂) | fn(item₃) | fn(item₄) | fn(item₅) |
```

List Expressions

output_list = [ fn(item) for item in input_iterable ]

1. Input Iterable
2. Temporary Iteration Variable
3. Iteration expression on temp variable
List Expressions

output_list = [fn(item) for item in input_iterable]

1. Input iterable: output list will match the number of items in this iterable

2. A temporary iteration variable: You can give it any name, but make it short but descriptive. Its type is the type of the individual members of the input iterable

3. The iteration computation that will transform each item in the input iterable to its corresponding item in the output list
List Expressions

```python
numbers = range(10)
squares = [i**2 for i in numbers]
```

<table>
<thead>
<tr>
<th>numbers</th>
<th>squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>9</td>
<td>81</td>
</tr>
</tbody>
</table>
words = "List expressions are cool!".split()
word_lengths = [len(word) for word in words]
List Expressions

output_list = [fn(item) for item in input_iterable]
Your Homework

• Implement two readability algorithms. One is simple and requires very little iteration. The second requires more complicated iteration.

• It requires you to break a text into 100 word chunks, which is a list of lists. You’ll learn how to do this in your activity.

• To perform computation over a list of lists, you will be creating nested list expressions.
Average number of sentences per 100 words

Average sentences per 100 words = \( \frac{\text{number sentences}}{\text{number chunks}} \)

- While these averages can be computed without list expressions using some Python built-in functions, for this assignment we’re requiring you to use list-expressions
- Your program will first create a list of lists, the 100-word chunks
Nested Iteration

- List of word lists
  - Word chunks
- Nested For Loop
- List of sentence counts
  - words_chunks
  - sentence_counts
  - List Expressions are cool!
Nested Iteration
Nested Iteration

1. Input iterable: a list of word lists, `word_chunks`
2. Iteration Variable: a single list of words
3. Iteration computation: A function that counts sentences in a single word list. You will write this function in the class activity
4. Output list: a list of integers, the sentence counts

```python
#(4) = [(3) for (2) in (1)]
```
Sentence count

Word list

word_1 word_2 word_3 word_4 ...

Summarizing function?

Sentence count 5
Sentence count

word list: \( \text{word}_1 \quad \text{word}_2 \quad \text{word}_3 \quad \text{word}_4 \quad \ldots \)

sentence-ending punctuation counts:

\( \begin{array}{cccc} 0 & 1 & 0 & 0 \end{array} \ldots \)

Summarizing function:

sentence count
Sentence count

word list

sentence-ending punctuation counts

sentence count

sum()
Conclusions

• Iteration enables us to re-use code an arbitrary number of times

• List expressions, one type of iteration, will produce a list the same size as the input iterable, which each element transformed by the specified expression

• For list expressions, you can forget the colon as they are not needed