

# CSE 2050 – Programming in a Second Language

Fall 2023

## Homework 3: Python Programming with Functions

Total Points: 30

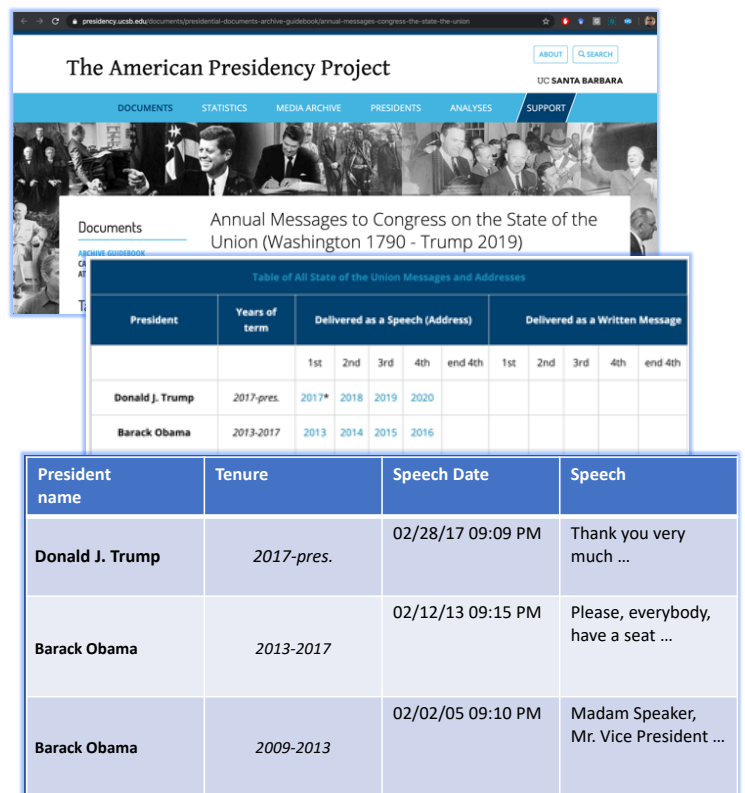
Date Assigned: Friday, Sept 22, 2023

Due Date: Sunday, Oct 1, 2023

**Instructions:** Please use functions as necessary throughout the assignment. Submit your work on Canvas as a Jupyter Notebook ipynb file named `cse2050_your_group_number_hw2.ipynb`. Remember to only use concepts covered up to Week 5 of classes to solve the problems. For example, do not use regular expressions, packages outside of the Python standard library, or other concepts not covered in class. See the *Tips for Success* and *Course Requirements* for more information.

### Key Programming Concepts Demonstrated

- Web Scraping
  - Scraping textual data from a webpage (Q1)
- Programming experience
  - Structuring code and maintaining program flow (Q1)
  - Designing functions that work together to solve a problem (Q2)
- Working with data
  - Manipulating tabular data with different fields (Q1-Q3)
  - Formatting 2-dimensional data (Q1-Q3)
  - Classifying article readability and grade level based on Flesch Index and Flesch-Kincaid Grade Level Formula (Q2)
- Bonus (Visualization)
  - Creating a bar chart from data in a table (Q3)



The screenshot shows the website 'The American Presidency Project' with a navigation bar including DOCUMENTS, STATISTICS, MEDIA ARCHIVE, PRESIDENTS, ANALYSES, and SUPPORT. The main content area is titled 'Documents' and 'Annual Messages to Congress on the State of the Union (Washington 1790 - Trump 2019)'. Below this is a table titled 'Table of All State of the Union Messages and Addresses'.

President	Years of term	Delivered as a Speech (Address)					Delivered as a Written Message				
		1st	2nd	3rd	4th	end 4th	1st	2nd	3rd	4th	end 4th
Donald J. Trump	2017-pres.	2017*	2018	2019	2020						
Barack Obama	2013-2017	2013	2014	2015	2016						

President name	Tenure	Speech Date	Speech
Donald J. Trump	2017-pres.	02/28/17 09:09 PM	Thank you very much ...
Barack Obama	2013-2017	02/12/13 09:15 PM	Please, everybody, have a seat ...
Barack Obama	2009-2013	02/02/05 09:10 PM	Madam Speaker, Mr. Vice President ...

Recall the web scraping activities demonstrated in class during the lectures from week 4-5. This assignment will build on some of that code. Particularly, we will scrape a collection of U.S. presidential State of the Union addresses from the web and use the Flesch Index to classify their readability according to grade levels.

### What is Flesch Index?

In 1949, Dr. Rudolf Flesch published *The Art of Readable Writing*, in which he proposed a measure of text readability known as the Flesch Index. This index is based on the average number of syllables per word and the average number of words per sentence in a piece of text. Index scores usually range from 0 to 100, and indicate readable prose for the following grade levels:

Score	School level	Notes
100.00–90.00	5th grade	Very easy to read. Easily understood by an average 11-year-old student.
90.0–80.0	6th grade	Easy to read. Conversational English for consumers.
80.0–70.0	7th grade	Fairly easy to read.
70.0–60.0	8th & 9th grade	Plain English. Easily understood by 13- to 15-year-old students.
60.0–50.0	10th to 12th grade	Fairly difficult to read.
50.0–30.0	College	Difficult to read.
30.0–0.0	College graduate	Very difficult to read. Best understood by university graduates.

Table 1: Article Readability based on Flesch Index

Flesch's formula to calculate the index  $F$  is the following:

$$F = 206.835 - 1.015 \times (\text{words.count} / \text{sentences.count}) - 84.6 \times (\text{syllables.count} / \text{words.count})$$

The Flesch-Kincaid Grade Level Formula is used to compute the Equivalent Grade Level  $G$ :

$$G = 0.39 \times (\text{words.count} / \text{sentences.count}) + 11.8 \times (\text{syllables.count} / \text{words.count}) - 15.59$$

Definitions	
Word:	Any sequence of non-whitespace characters.
Sentence:	Any sequence of words ending in a period, question mark, exclamation point, colon, or semicolon.
Syllable:	Any word of three characters or less; or any vowel (a, e, i, o, u) or pair of consecutive vowels, except for a final -es, -ed, or -e that is not -le

Table 2: Definition of terms related to the Flesch Index

First, access the code located at <https://repl.it/@fnembhard/souspeechscraper>. Copy the code to your Jupyter notebook. This code scrapes data from the webpage of [The American Presidency Project at UC Santa Barbara](#) and extracts links to State of the Union (SOU) addresses. The code only selects the first SOU delivered by each president during each term if available. Further, the code prints out a table containing the following fields for each president: [president\_name, tenure, speech\_link, speech\_date, speech]. (speech\_date and speech are left for you to update using Python code).

#### 1. Get the Data (10 points)

Update the code by adding a function named `get_speech_and_speech_date(url)` to extract each speech and speech date from the page referenced by the link in the table and update your table (2-D list). Remember you can return multiple values from a function as a tuple. This can be done by accessing the object containing the desired text by `class`. Figure 1 shows the layout of the page for each speech. For example, to extract the element containing the speech date, use the following code:

```
speech_date_elmnt = html_elem.cssselect('[class="date-display-single"]')
```

Figure 1: The American Presidency Project SOU page format

```

182
183
184 <div class="field-ds-doc-title">
185 <h1>Address Before a Joint Session of the Congress</h1>
186 </div>
187
188 </div>
189
190
191
192
193
194
195
196
197
198
199
200 <div class="field-docs-start-date-time">
201 <span class="date-display-single" property="dc:date" datatype="xsd:dateTime" content="2017-02-28T21:09:00+00:00">February 28, 2017</span> </div>
202
203
204
205 <div class="field-docs-content">
206 <p>Thank you very much. Mr. Speaker, Mr. Vice President, Members of Congress, the First Lady of the United States, and citizens of America: Tonight, as we mark the conclusion of our celebration of Black History Month, we are reminded of our Nation's path towards civil rights and the work that still remains to be done. Recent threats targeting Jewish community centers and vandalism of Jewish cemeteries, as well as last week's shooting in Kansas City, remind us that while we may be a nation divided on policies, we are a country that stands united in condemning hate and evil in all of its very ugly forms.
207 </p><p>Each American generation passes the torch of truth, liberty, and justice in an unbroken chain, all the way down to the present. That torch is now in our hands, and we will use it to light up the world. I am here tonight to deliver a message of unity and strength, and it is a message deeply delivered from my heart. A new chapter of American greatness is now beginning. A new national pride is sweeping across our Nation. And a new surge of optimism is placing impossible dreams firmly within our grasp.
208 </p><p>What we are witnessing today is the renewal of the American spirit. Our allies will find that America is once again ready to lead. All the nations of the world—friend or foe—will find that America is strong, America is proud, and America is free.
209 </p><p>In 9 years, the United States will celebrate the 250th anniversary of our founding: 250 years since the day we declared our independence. It will be one of the great milestones in the history of the world. But what will America look like as we reach our 250th year? What kind of country will we leave for our children?
210 </p><p>I will not allow the mistakes of recent decades past to define the course of our future. For too long, we've watched our middle class shrink as we've exported our jobs and wealth to foreign countries. We've financed and built one global project after another, but ignored the fates of our children in the inner cities of Chicago, Baltimore, Detroit, and so many other places throughout our land.
211 </p><p>We've defended the borders of other nations, while leaving our own borders wide open for anyone to cross and for drugs to pour in at a now unprecedented rate. And we've spent trillions and trillions of dollars overseas, while our infrastructure at home has so badly crumbled.
212 </p><p>Then, in 2016, the Earth shifted beneath our feet. The rebellion started as a quiet protest, spoken by families of all colors and creeds, families who just wanted a fair shot for their children and a fair hearing for their concerns.
213 </p><p>But then the quiet voices became a loud chorus, as thousands of citizens now spoke out together, from cities small and large, all across our country. Finally, the chorus became an earthquake, and the people turned out by the tens of millions, and they were all united by one very simple, but crucial demand: that America must put its own citizens first. Because only then can we truly make America great again.
214 </p><p>Dying industries will come roaring back to life. Heroic veterans will get the care they so desperately need. Our military will be given the resources its brave warriors so richly deserve. Crumbling infrastructure will be replaced with new roads, bridges, tunnels, airports, and railways gleaming across our very, very beautiful land. Our terrible drug epidemic will slow down and, ultimately, stop. And our neglected inner cities will see a rebirth of hope, safety, and opportunity. Above all else, we will keep our promises to the American people. [<i>Applause</i>] Thank you.
215 </p><p>It's been a little over a month since my Inauguration, and I want to take this moment to update the Nation on the progress I've made in keeping those promises.

```

Use the *content* attribute to extract the actual speech date (full date with time), which is the actual date the speech was delivered. Next, extract the speech transcript and store it in your table. Print out a table of your results. Only print the first 50 characters of each speech you extract when displaying the output. Please pay attention to the reformatted and more readable date in the table.

## Expected Output

President's Name	Term	Date of Speech	Excerpt
Joseph R. Biden	2021-present	04/28/21 09:06 PM	The President. Thank you. Thank you. Thank you. Go
Donald J. Trump	2017-pres.	02/28/17 09:09 PM	Thank you very much. Mr. Speaker, Mr. Vice Preside
Barack Obama	2013-2017	02/12/13 09:15 PM	Please, everybody, have a seat. Mr. Speaker, Mr. V
George W. Bush	2005-2009	02/02/05 09:10 PM	Mr. Speaker, Vice President Cheney, Members of Con
William J. Clinton	1997-2001	02/04/97 09:15 PM	Mr. Speaker, Mr. Vice President, Members of the 10
George Bush	1989-1993	02/09/89 12:00 AM	Mr. Speaker, Mr. President, and distinguished Memb
Ronald Reagan	1985-1989	02/06/85 09:05 PM	Mr. Speaker, Mr. President, distinguished Members
Lyndon B. Johnson	1965-1969	01/04/65 09:04 PM	Mr. Speaker, Mr. President, Members of the Congres
John F. Kennedy	1961-1963	01/30/61 12:30 PM	[As delivered in person before a joint session]Mr.
Dwight D. Eisenhower	1957-1961	01/10/57 12:00 AM	To the Congress of the United States:I appear befo
Harry S Truman	1949-1953	01/05/49 01:00 PM	Mr. President, Mr. Speaker, Members of the Congres
Franklin D. Roosevelt	1945	01/06/45 12:00 AM	Today, in pursuance of my Constitutional duty, I s
Warren G. Harding	1921-1923	12/06/21 12:00 AM	MR. SPEAKER AND MEMBERS OF THE CONGRESS:It is a ve
Woodrow Wilson	1917-1921	12/04/17 12:00 AM	Gentlemen of the Congress:Eight months have elapse
John Adams	1797-1801	11/22/97 12:00 AM	Gentlemen of the Senate and Gentlemen of the House
George Washington	1793-1797	12/03/93 12:00 AM	Fellow-Citizens of the Senate and House of Represe

## 2. Writing Functions (20 points)

Write code to create the following functions to find the number of sentences, words, and syllables in each speech, as well as the speech's Flesch Index and Grade Level Equivalent. Refer to the formulas previously defined and the definitions in Table 2. Update your table with the Flesch Index and grade level as shown in the expected output below.

**count\_sentences(speech):** Counts the number of sentences in text.

**count\_words(speech):** Counts the number of words in text.

**count\_syllables(speech):** Counts the number of syllables in text.

**compute\_flesch\_index(syllable\_count, word\_count, sentence\_count):** Computes the Flesch Index for the given numbers of sentences, words, and syllables.

**compute\_grade\_level(syllable\_count, word\_count, sentence\_count):** Computes the Grade Level Equivalent for the given numbers of sentences, words, and syllables.

**classify\_article\_readability(flesch\_index):** Returns the school/grade level in worded format based on Table 1.

(Hint: the sentence count is a count of all periods, question marks, exclamation marks, colons and semicolons. To count words, simply split the text by space. To count syllables, count the vowels in each word plus all words that end with "le". Subtract 1 from the syllable count if the word ends with 'es', 'ed', or 'e' (not 'le'))

## Expected Output

President's Name	Term	Date of Speech	Excerpt	Flesch	Grade Level #	Grade Level
Joseph R. Biden	2021-present	04/28/21 09:06 PM	The President. Thank you. Than	57.64	9	10th to 12th grade
Donald J. Trump	2017-pres.	02/28/17 09:09 PM	Thank you very much. Mr. Speak	53.97	10	10th to 12th grade
Barack Obama	2013-2017	02/12/13 09:15 PM	Please, everybody, have a seat	54.46	10	10th to 12th grade
George W. Bush	2005-2009	02/02/05 09:10 PM	Mr. Speaker, Vice President Ch	45.84	12	College
William J. Clinton	1997-2001	02/04/97 09:15 PM	Mr. Speaker, Mr. Vice Presiden	56.20	10	10th to 12th grade
George Bush	1989-1993	02/09/89 12:00 AM	Mr. Speaker, Mr. President, an	58.46	9	10th to 12th grade
Ronald Reagan	1985-1989	02/06/85 09:05 PM	Mr. Speaker, Mr. President, di	49.72	10	College
Lyndon B. Johnson	1965-1969	01/04/65 09:04 PM	Mr. Speaker, Mr. President, Me	53.14	10	10th to 12th grade
John F. Kennedy	1961-1963	01/30/61 12:30 PM	[As delivered in person before	36.56	14	College
Dwight D. Eisenhower	1957-1961	01/10/57 12:00 AM	To the Congress of the United	34.90	13	College
Harry S Truman	1949-1953	01/05/49 01:00 PM	Mr. President, Mr. Speaker, Me	42.21	12	College
Franklin D. Roosevelt	1945	01/06/45 12:00 AM	Today, in pursuance of my Cons	46.84	12	College
Warren G. Harding	1921-1923	12/06/21 12:00 AM	MR. SPEAKER AND MEMBERS OF THE	34.33	15	College
Woodrow Wilson	1917-1921	12/04/17 12:00 AM	Gentlemen of the Congress:Eigh	52.71	12	10th to 12th grade
John Adams	1797-1801	11/22/97 12:00 AM	Gentlemen of the Senate and Ge	30.95	17	College
George Washington	1793-1797	12/03/93 12:00 AM	Fellow-Citizens of the Senate	39.37	15	College

**Bonus (7 points)**

Research and use matplotlib to display a plot of the Flesch Indexes for all the presidents in our sample dataset as shown in the figure below. Do not hard-code any data. Extract data from your data-table and use it to create your plot.

Figure 2: Plot of Flesch indexes for a selection of SOU for various US presidents

