HW0 - Getting Started with Assignment reports and Linux command-line utilities

Courtney Maynard

September 8th, 2024(September 10th, 2024)

Q3

Replace the "Growth of the Early Web" image. Make sure to change the image caption and description of the image in the report, too.

Answer

Figure 1 shows me hiking in Acadia National Park this past summer.

Figure 1: Courtney hiking in Acadia National Park

Discussion

I arrived at the image shown above by following the template for uploading an image in Latex. The implication of this exercise is that I now know how to display images in Latex.

Q4

Replace the code in the lstlisting environment with another block of code. You can use some Python code, or you can insert code from a different language – just change the language indicated so that syntax highlighting still works properly. Make sure to change the caption as needed.

Answer

Listing 1 is an example of directly copying code into the LaTeX document and having the listings package perform syntax highlighting. Listing 2 is an example of importing the code from a file rather than copying it in.

```
# must split lyrics into lines, in order to retain the line-level information.
def lyrics_into_lines(lyrics):
    return lyrics.split('\n')

df_balanced['Lyrics_Lines'] = df_balanced['Lyrics'].apply(lyrics_into_lines)

# tokenize each line individually in order to respect the structure of the song.

def tokenize_line(line):
    line = re.sub(r'[^a-zA-Z0-9\s]', '', line.lower())
    tokens = word_tokenize(line)
    return tokens

def tokenize_lyrics(lines):
    return [tokenize_line(line) for line in lines]

df_balanced['Tokenized_Lyrics'] = df_balanced['Lyrics_Lines'].apply(tokenize_lyrics)
```

Listing 1: Tokenization of song lyrics copied into the LaTeX

```
# must split lyrics into lines, in order to retain the line-level information.
```

```
def lyrics_into_lines(lyrics):
    return lyrics.split('\n')

df_balanced['Lyrics_Lines'] = df_balanced['Lyrics'].apply(
    lyrics_into_lines)

# tokenize each line individually in order to respect the structure of
    the song.

def tokenize_line(line):
    line = re.sub(r'[^a-zA-Z0-9\s]', '', line.lower())
    tokens = word_tokenize(line)
    return tokens

def tokenize_lyrics(lines):
    return [tokenize_line(line) for line in lines]

df_balanced['Tokenized_Lyrics'] = df_balanced['Lyrics_Lines'].apply(
    tokenize_lyrics)
```

Listing 2: Tokenization of song lyrics loaded from file

Discussion

I followed the template to figure out how to display python code in the two different ways: through directly pasting it in and through uploading a file. As a result of this exercise, I now know how to display code in Latex using two different methods.

Q5

Edit Table 1 so that it matches the first 4 weeks of our class schedule, as given in our syllabus.

Answer

Table 1 shows our class syllabus for the first four weeks.

Table 1: First Four Weeks Schedule

Week	Lecture Dates	Topic	Homework (Date Assigned – Due Date
01	Aug 29 & Sep 3	Introduction to Web Science and Web Architecture	HW0 - Getting started, Aug 29 - - Sep 10
02	Sep 5 & 10	Introduction to Python	Python Google Colab notebook Python lab exercises HW1 - Web Sci. Intro, Sep 10 – 24
03	Sep 12 & 17	Introduction to Info Vis with R, Python Web Scraping	InfoVis in R Colab notebook InfoVis in Python Colab notebook Web Scraping (IMDB) Python Colab notebook Web Scraping (Twitter) Python scripts
04	Sep 19 & 24	Measuring the Web	

Discussion

I used the example table to begin creating this table. In order to make sure that the different homework links were able to be stacked inside their respective columns, I looked at Stackexchange to learn how to format the table correctly. Thus, I changed the column types in order to use the newline command to create the corresponding table. Additionally, I linked all of the respective lectures and homework using the href command, which I knew how to do because of previous Latex experience. Because of this exercise, I now know how to format various types of tables in Latex.

References

- Stackexchange - Latex - Table with multiple lines in some cells, `https://tex.stackexchange.com/questions/40561/table-with-multiple-lines-in-some-cells`
- Scribbr - Poisson Distributions — Definition, Formula & Examples, `https://www.scribbr.com/statistics/poisson-distribution/`
- Computer Science Wiki - Graph theory and connectivity of the web, `https://computersciencewiki.org/index.php/Graph_theory_and_connectivity_of_the_web`