LifeWeb GuidedLDA Topic Tagging Guide

1. Ensure Anaconda is installed on your machine

- In order to open the .ipynb files you must use Jupyter Notebook, we recommend accessing this through Anaconda's Navigator
    - Install Anaconda on your machine: https://docs.anaconda.com/anaconda/install/
    - Download the Anaconda Navigator by running
        ```
        conda install -c anaconda anaconda-navigator
        ```
        in your terminal.
    - Launch Anaconda Navigator, launch Jupyter Notebook from it (you'll be redirected to a new tab in your browser running Jupyter Notebook).
    - Navigate to the relevant repo, and open/manipulate files in your browser!

For a very comprehensive introduction to Jupyter Notebook, click here.

2. Install packages

- For webscraping
    - Selenium:
    - Beautiful Soup
        ```
        pip install beautifulsoup4
        ```
- For guidedLDA
    - Explanation:
      https://www.freecodecamp.org/news/how-we-changed-unsupervised-lda-to-semi-supervised-guidedlda-e36a95f3a164/
    - Follow the installation guide of this git page:
      https://github.com/vi3k6i5/GuidedLDA
    - Spoiler Alert: it's essentially just
        ```
        pip install guidedlda
        ```

3. Python Notebooks/Code

- Webscraping
    - eecs338_webscraping.ipynb
- Processing text (if necessary; i.e. lemmatize,)
    - eecs338_data_preprocessing.ipynb (script - not necessary)
        - generates "preprocessed_text.txt"
    - forevermissed_data_cleaning.py
        - Generates "lemma_forevermissed_story_final1.txt" and "lemma_forevermissed_tribute_final.txt"

- GuidedLDA
    - Forevermissed_guidedlda.py
        - Run .py files with `python` and then the file's name
    - eecs338_test.ipynb (script?)

Mockup
- Edit mode: https://edit.genial.ly/5cf082b34a64dd0f335a330d
    - Used Gravit to design, can use any graphic design tool
- Final: https://view.genial.ly/5cf082b34a64dd0f335a330d/interactive-content-lifeweb-mockup