

Natural Language Parser and Taxonomy Creation Tool Startup Guide



Brandon Christensen, Kadir Nour, Riley Hunter

4/28/2022

Table of Contents

Installation.....	3
Setup.....	6
Running the Program.....	9
Service Walkthrough.....	10

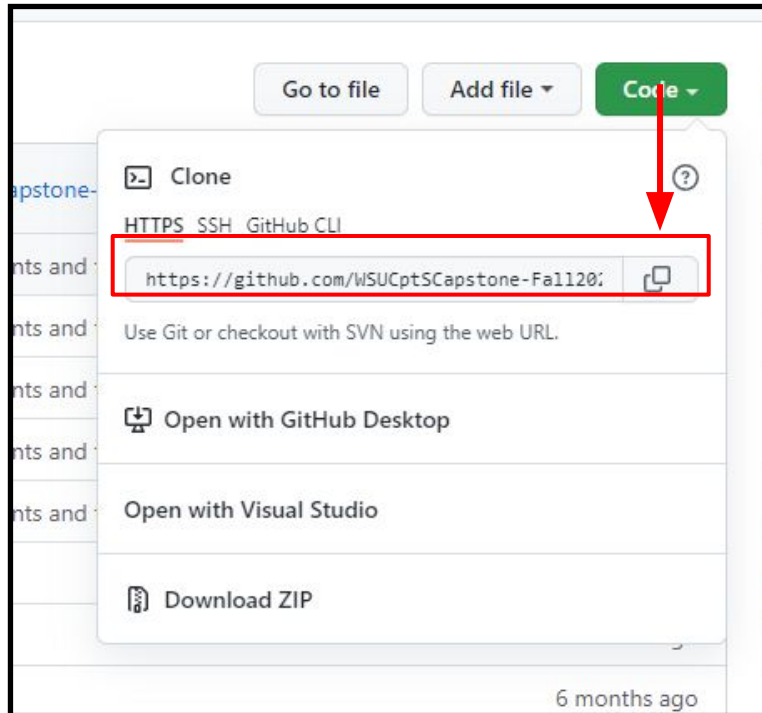
Installation

Downloading the Project

1. Go to our project page on Github:

<https://github.com/WSUCptSCapstone-Fall2021Spring2022/boeing-naturallanguageprocessing>

a. Copy the projects' HTTPS URL.



2. Open a terminal in the directory you want to download the project in:

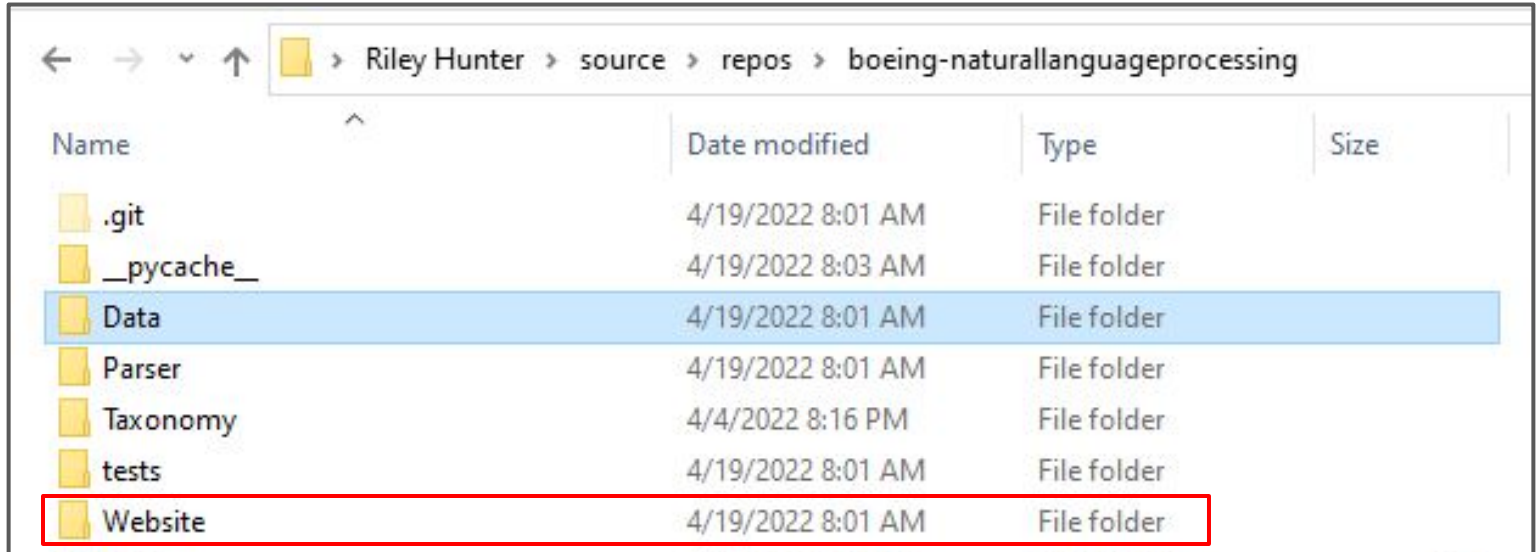
a. Enter "git clone <our repo>".

```
MINGW64: c:/Users/blcsi/OneDrive/Desktop
blcsi@BLD MINGW64 ~/OneDrive/Desktop
$ git clone https://github.com/WSUCptSCapstone-Fall2021Spring2022/boeing-naturallanguageprocessing
```

Installation

Installing Dependencies

3. Open Website folder



Name	Date modified	Type	Size
.git	4/19/2022 8:01 AM	File folder	
__pycache__	4/19/2022 8:03 AM	File folder	
Data	4/19/2022 8:01 AM	File folder	
Parser	4/19/2022 8:01 AM	File folder	
Taxonomy	4/4/2022 8:16 PM	File folder	
tests	4/19/2022 8:01 AM	File folder	
Website	4/19/2022 8:01 AM	File folder	

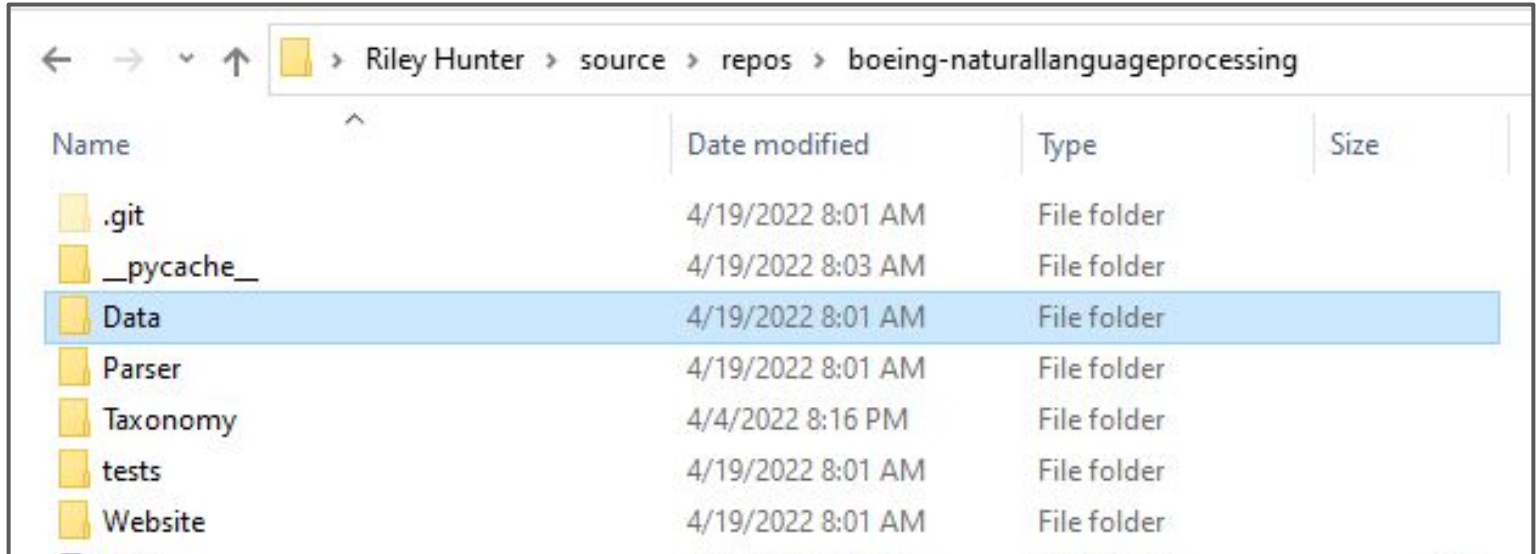
a. Open a terminal and enter “npm install”

```
riley@MSI MINGW64 ~/source/repos/boeing-naturallanguageprocessing/Website (main)
$ npm install
npm WARN deprecated flatten@1.0.3: flatten is deprecated in favor of utility fra
npm WARN deprecated @hapi/topo@3.1.6: This version has been deprecated and is no
npm WARN deprecated @hapi/bourne@1.3.2: This version has been deprecated and is
npm WARN deprecated urix@0.1.0: Please see https://github.com/lydell/urix#deprec
npm WARN deprecated chokidar@2.1.8: Chokidar 2 will break on node v14+. Upgrade
```

Installation

Installing Dependencies

2. Open boeing-naturallanguageprocessing (project) folder



a. Open a terminal and enter “pip install -r requirements.txt”

```
(.venv) C:\Users\User\Documents\Cap\boeing-naturallanguageprocessing-fork [main ± +0 ~7 -0 !]> pip install -r .\requirements.txt
Requirement already satisfied: atomicwrites==1.4.0 in c:\users\user\documents\cap\.venv\lib\site-packages (from -r .\requirements.txt (line 1)) (1.4.0)
Requirement already satisfied: attrs==20.3.0 in c:\users\user\documents\cap\.venv\lib\site-packages (from -r .\requirements.txt (line 2)) (20.3.0)
Requirement already satisfied: autopep8==1.5.7 in c:\users\user\documents\cap\.venv\lib\site-packages (from -r .\requirements.txt (line 3)) (1.5.7)
Requirement already satisfied: blis==0.7.4 in c:\users\user\documents\cap\.venv\lib\site-packages (from -r .\requirements.txt (line 4)) (0.7.4)
Requirement already satisfied: catalogue==2.0.1 in c:\users\user\documents\cap\.venv\lib\site-packages (from -r .\requirements.txt (line 5)) (2.0.1)
Requirement already satisfied: certifi==2020.12.5 in c:\users\user\documents\cap\.venv\lib\site-packages (from -r .\requirements.txt (line 6)) (2020.12.5)
```

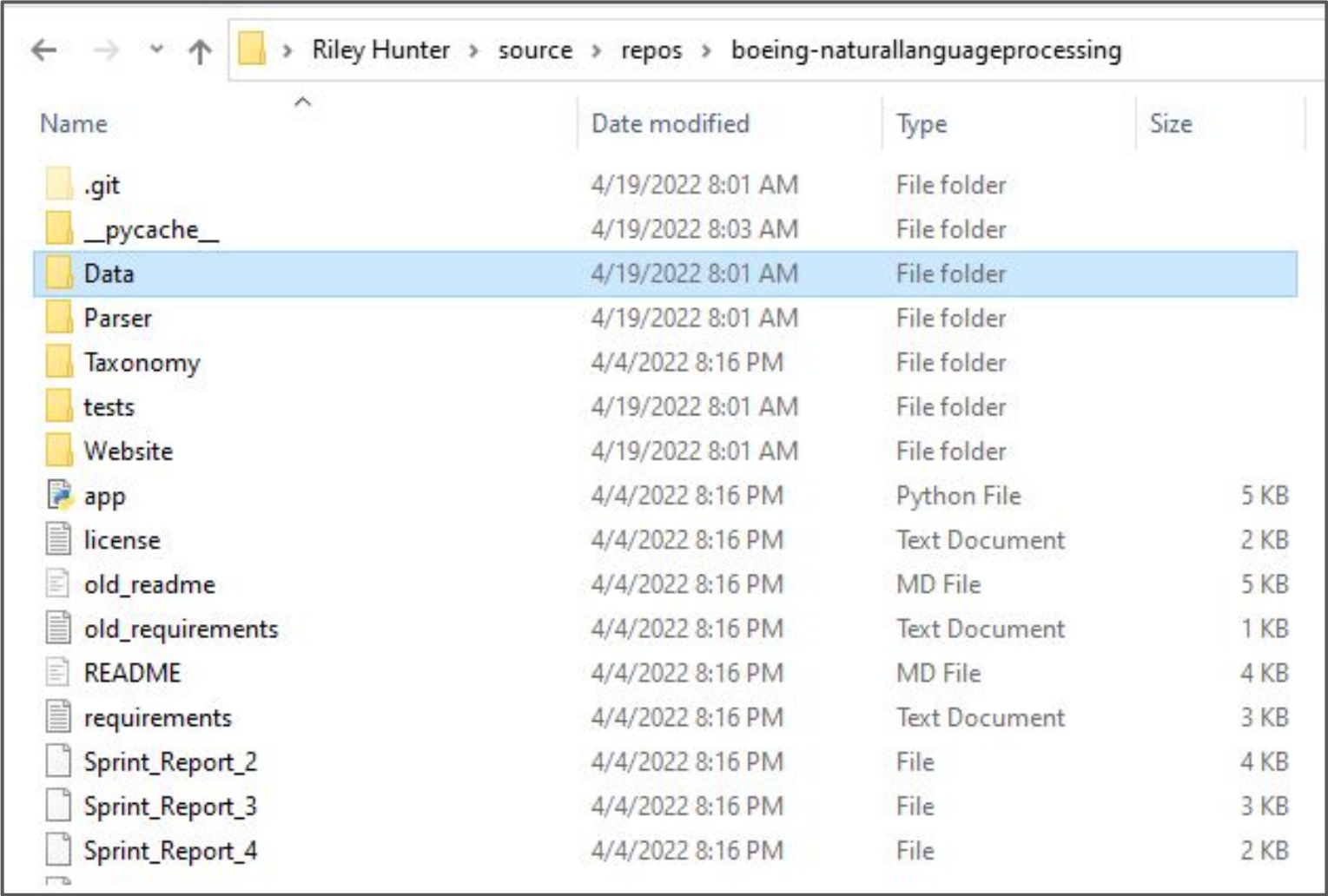
b. In the same terminal also enter “python -m spacy download en_core_web_sm”

```
.8.2)
Requirement already satisfied: Werkzeug==1.0.1 in c:\users\user\documents\cap\.venv\lib\site-packages (from -r .\requirements.txt (line 59)) (1.0.1)
Requirement already satisfied: setuptools in c:\users\user\documents\cap\.venv\lib\site-packages (from spacy==3.0.3->-r .\requirements.txt (line 49)) (41.2.0)
(.venv) C:\Users\User\Documents\Cap\boeing-naturallanguageprocessing-fork [main ± +0 ~7 -0 !]> python -m spacy download en_core_web_sm
Collecting en-core-web-sm==3.0.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.0.0/en_core_web_sm-3.0.0-py3-none-any.whl (13.7 MB)
```

Setup

Uploading Documents

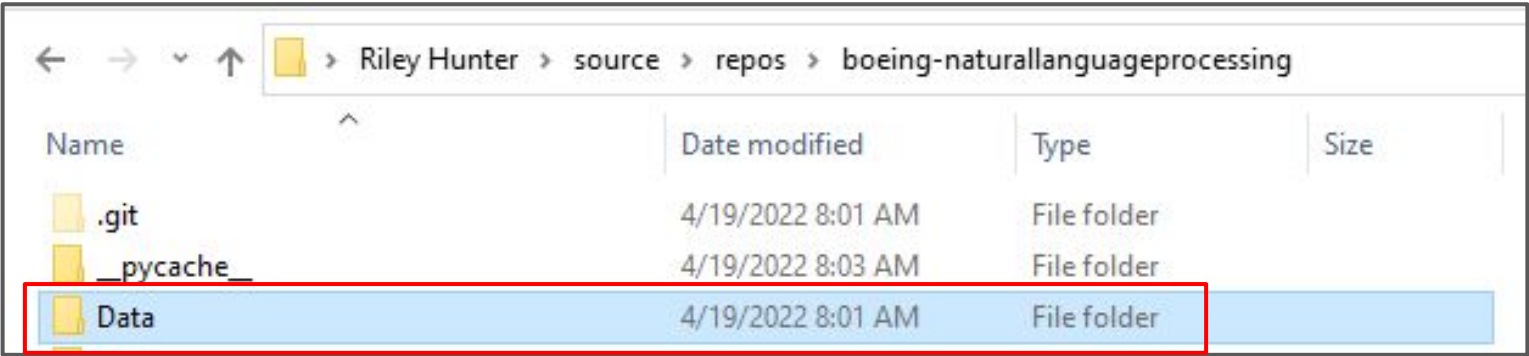
1. Open boeing-naturallanguageprocessing (project) folder:



The screenshot shows a file explorer window with the address bar set to 'Riley Hunter > source > repos > boeing-naturallanguageprocessing'. The table below lists the contents of this folder.

Name	Date modified	Type	Size
.git	4/19/2022 8:01 AM	File folder	
__pycache__	4/19/2022 8:03 AM	File folder	
Data	4/19/2022 8:01 AM	File folder	
Parser	4/19/2022 8:01 AM	File folder	
Taxonomy	4/4/2022 8:16 PM	File folder	
tests	4/19/2022 8:01 AM	File folder	
Website	4/19/2022 8:01 AM	File folder	
app	4/4/2022 8:16 PM	Python File	5 KB
license	4/4/2022 8:16 PM	Text Document	2 KB
old_readme	4/4/2022 8:16 PM	MD File	5 KB
old_requirements	4/4/2022 8:16 PM	Text Document	1 KB
README	4/4/2022 8:16 PM	MD File	4 KB
requirements	4/4/2022 8:16 PM	Text Document	3 KB
Sprint_Report_2	4/4/2022 8:16 PM	File	4 KB
Sprint_Report_3	4/4/2022 8:16 PM	File	3 KB
Sprint_Report_4	4/4/2022 8:16 PM	File	2 KB

2. Open Data folder:



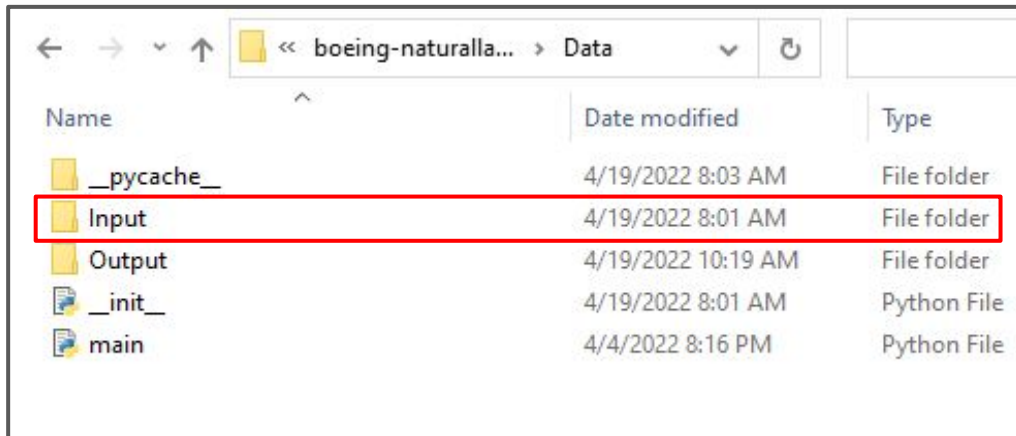
The screenshot shows the same file explorer window, but now the 'Data' folder is highlighted with a red box, indicating it is the selected folder to open.

Name	Date modified	Type	Size
.git	4/19/2022 8:01 AM	File folder	
__pycache__	4/19/2022 8:03 AM	File folder	
Data	4/19/2022 8:01 AM	File folder	

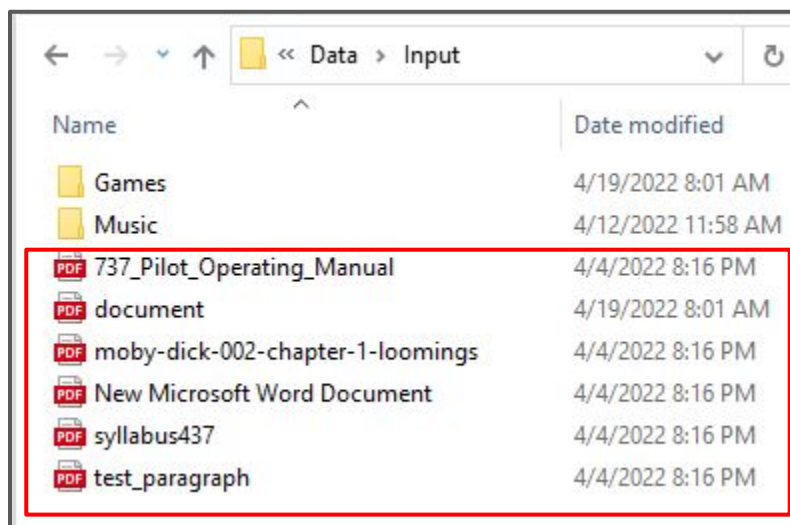
Setup

Uploading Documents

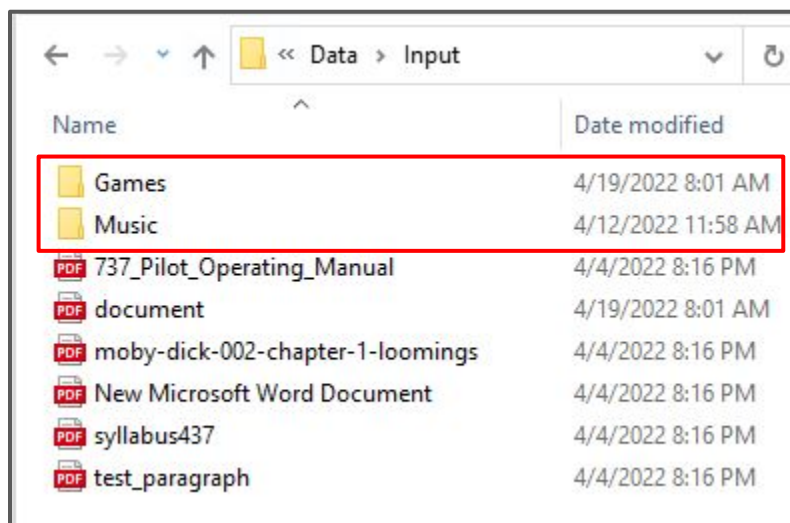
3. Open Input Folder:



a. Add any documents you want to parse (.txt, .docx, .pdf)



4. You can also make subfolders and add documents there:

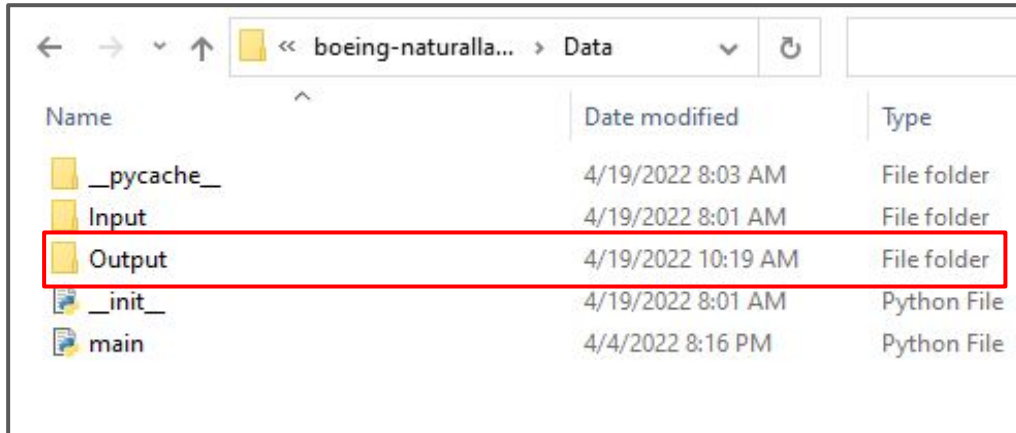


Setup

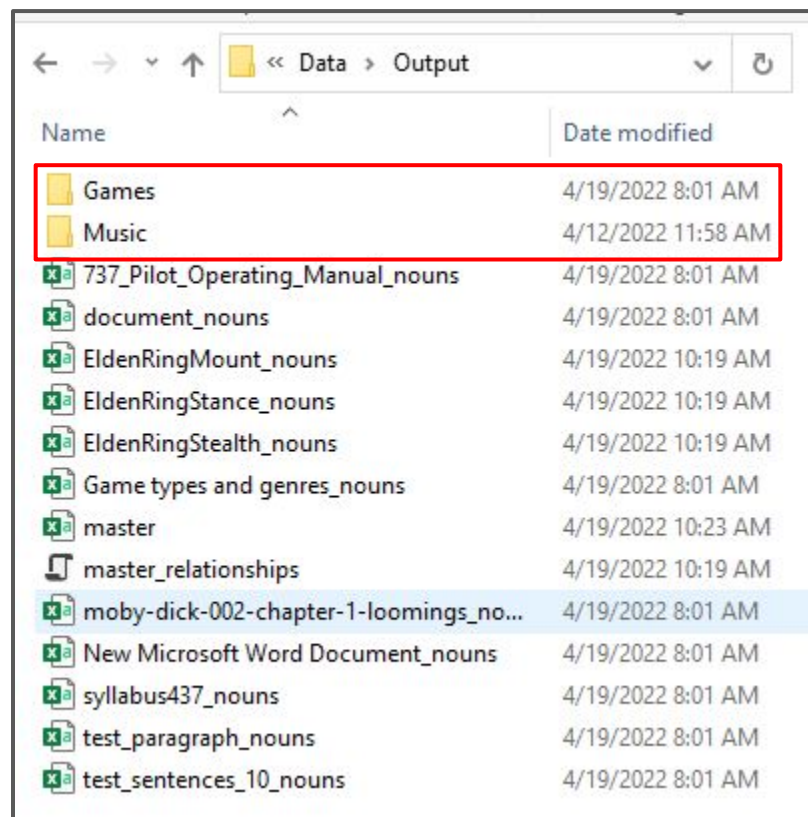
Selecting Output Location

1. Open Data folder:

a. Open Output folder



2. If you want to make subfolders for the output, do it here:



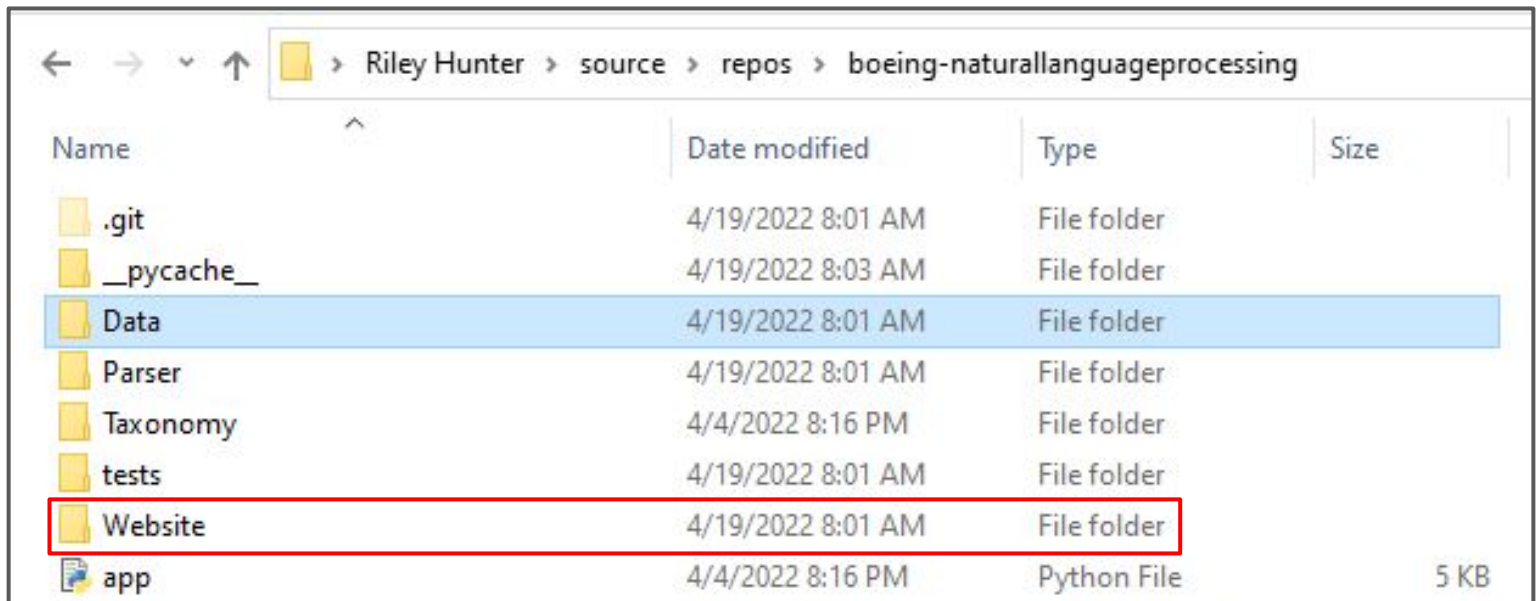
Running the Program

Starting the Servers

1. Open boeing-naturallanguageprocessing (project) folder
 - a. Open a terminal and enter “python -m flask run” (depending on python version)

```
riley@MST-MINGW64 ~/source/repos/boeing-naturallanguageprocessing (testing)
$ python3 -m flask run
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: off
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

2. Open Website folder



The screenshot shows a file explorer window with the path `Riley Hunter > source > repos > boeing-naturallanguageprocessing`. The following table represents the contents of this directory:

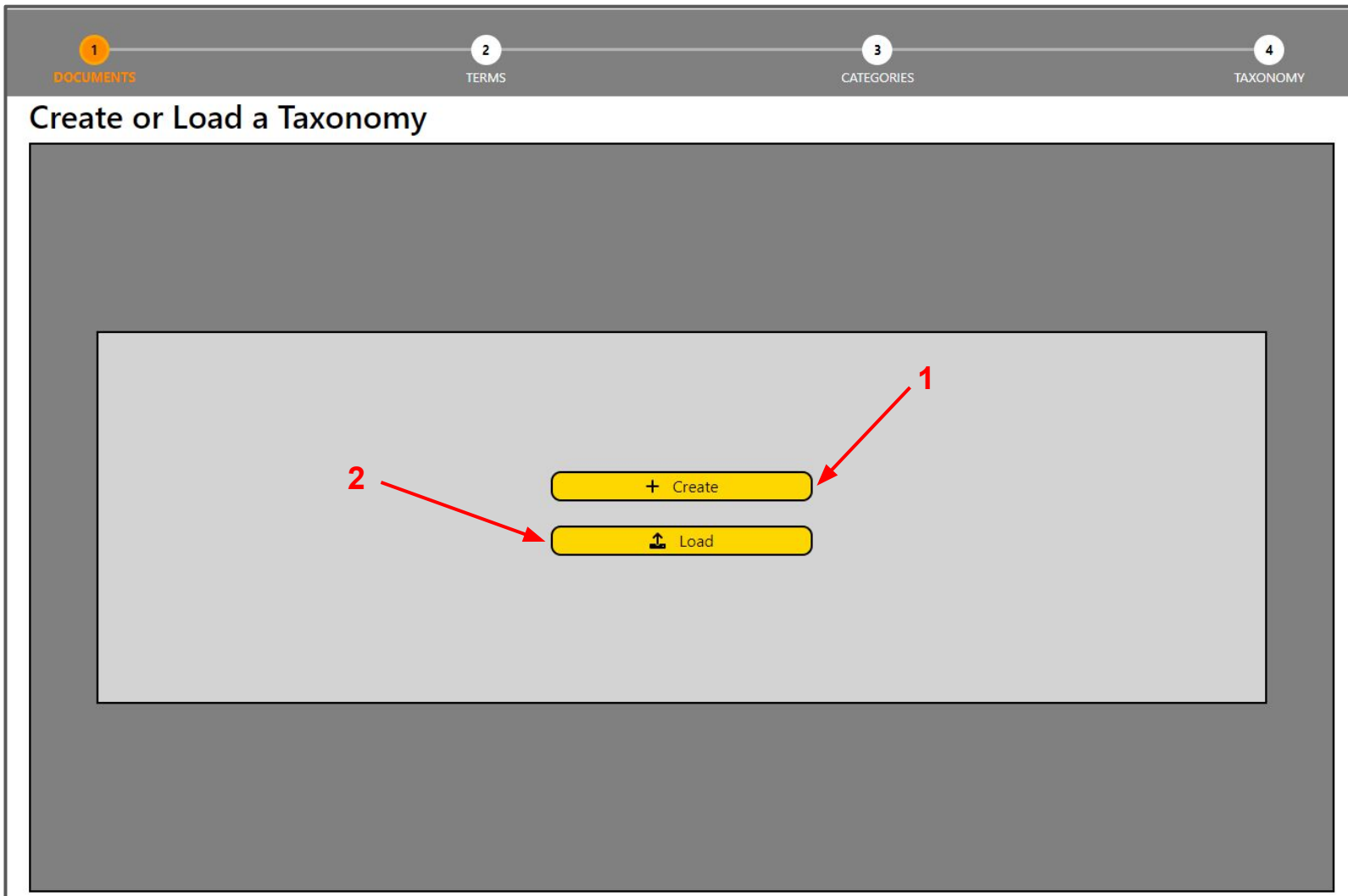
Name	Date modified	Type	Size
.git	4/19/2022 8:01 AM	File folder	
__pycache__	4/19/2022 8:03 AM	File folder	
Data	4/19/2022 8:01 AM	File folder	
Parser	4/19/2022 8:01 AM	File folder	
Taxonomy	4/4/2022 8:16 PM	File folder	
tests	4/19/2022 8:01 AM	File folder	
Website	4/19/2022 8:01 AM	File folder	
app	4/4/2022 8:16 PM	Python File	5 KB

- a. Open a terminal and enter “npm start”

```
riley@MST-MINGW64 ~/source/repos/boeing-naturallanguageprocessing/Website (testing)
$ npm start
```

Service Walkthrough

Create or Load a Taxonomy



1. Button to create a new taxonomy
2. Button to load a taxonomy

Service Walkthrough

Creating a New Taxonomy

The screenshot shows the 'Document Corpus (Create)' interface. At the top, there is a progress bar with four steps: 1. DOCUMENTS, 2. TERMS, 3. CATEGORIES, and 4. TAXONOMY. The interface is divided into several sections. The top section contains three text input boxes for 'Input', 'Output', and 'Name', each with a corresponding 'Submit' button. The 'Input' box is annotated with a red arrow and the number 1. The 'Output' box is annotated with a red arrow and the number 2. The 'Name' box is annotated with a red arrow and the number 3. The 'Submit Input Location' button is annotated with a red arrow and the number 4. The 'Submit Output Location' button is annotated with a red arrow and the number 5. The 'Submit Taxonomy Name' button is annotated with a red arrow and the number 6. Below these input boxes is a 'Get Files From Input Folder' button, annotated with a red arrow and the number 7. The bottom section is a table with two columns: 'File' and 'Add/ Remove'. The table lists several files, including '737_Pilot_Operating_Manual.pdf', 'Futball', 'Games', 'Music', 'New Microsoft Word Document.pdf', and 'moby-dick-002-chapter-1-loomings.pdf'. The 'Add/ Remove' column contains 'Delete' and 'Add' buttons for each file. The 'Delete' buttons are annotated with a red arrow and the number 8. The 'Add' buttons are annotated with a red arrow and the number 9. At the bottom of the interface, there are 'Back' and 'Forward' buttons, both annotated with a red arrow and the number 10.

1. Text input box to select an input directory

2. Text input box to select output directory

3. Text input box to select taxonomy name

4. Buttons to submit text input boxes

5. Button to load in documents from input directory

6. Buttons to add or remove documents to parse

7. Buttons to go forward and backward to next steps in taxonomy creation process

Service Walkthrough

Loading a Taxonomy

1DOCUMENTS

2TERMS

3CATEGORIES

4TAXONOMY

Document Corpus (Load)

1 → C:\Users\user\OneDrive\Desktop\boeing-naturallanguageprocessing\Data\Output Submit Output Location

Output:

2 → Will write over files with the same name Submit Taxonomy Name:

Name: master

Back

Forward

- 1. Text input box to select output directory (where the taxonomy is saved)
- 2. Text input to select the taxonomy to load

Service Walkthrough

Term Extraction

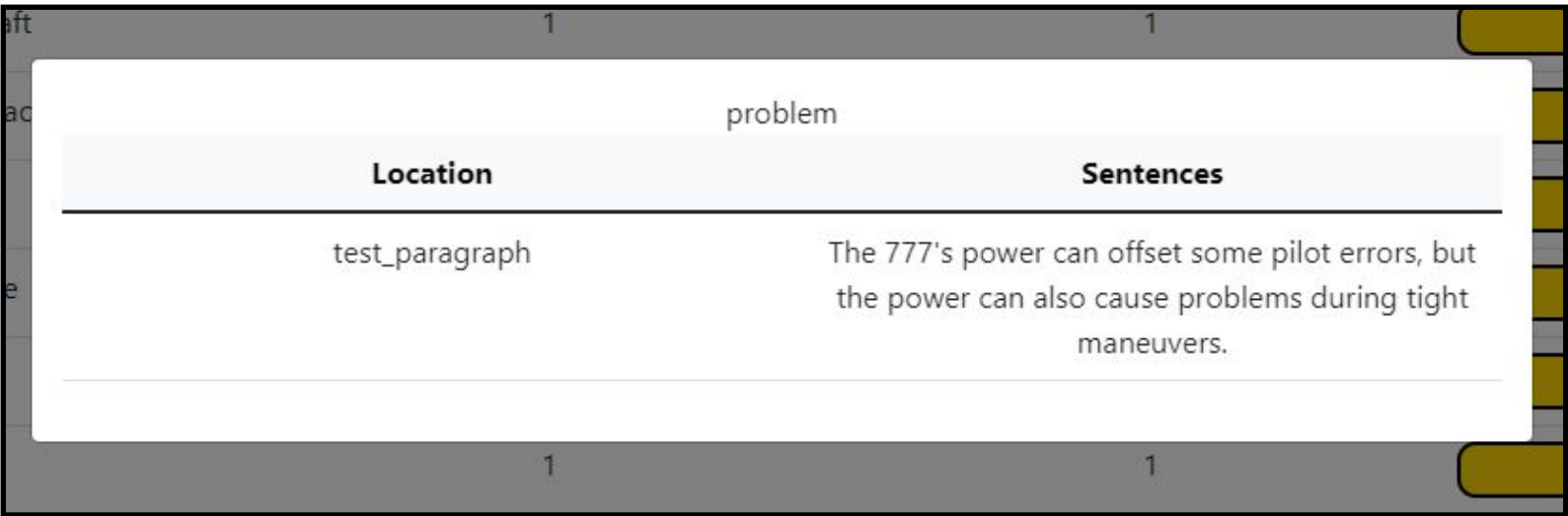
The screenshot shows the 'Term Extraction' interface. At the top, a navigation bar has four tabs: 1 DOCUMENTS, 2 TERMS, 3 CATEGORIES, and 4 TAXONOMY. The 'TERMS' tab is active. Below the navigation bar, the title 'Term Extraction' is displayed. The main area contains a table with columns: Noun, Frequency, Weight, and Sentences. The table lists various terms, some of which are highlighted in blue. To the right of the table, there is a 'Select terms to remove' section with a list of sentences. At the bottom, there are several buttons: 'Refresh Terms:', 'Save Terms:', '< Previous: 0', 'Next: >', 'Clear Selected', 'Delete Terms', 'Back', and 'Forward >>'. Red arrows and numbers 1 through 7 point to specific elements: 1 points to 'Refresh Terms:', 2 points to 'Save Terms:', 3 points to the 'Noun' column, 4 points to the 'Sentences' column, 5 points to the pagination controls, 6 points to 'Clear Selected', and 7 points to 'Delete Terms'.

Noun	Frequency	Weight	Sentences
a 737 or other small aircraft	1	1	Sentences
a very large aircraft	1	1	Sentences
a very short approach	1	1	Sentences
instance	1	1	Sentences
large aircraft rule	1	1	Sentences
pilot	2	2	Sentences
problem	1	1	Sentences
some pilot error	1	1	Sentences
the 777's power	1	1	Sentences
the most technologically advanced commercial aircraft	1	1	Sentences
the power	1	1	Sentences

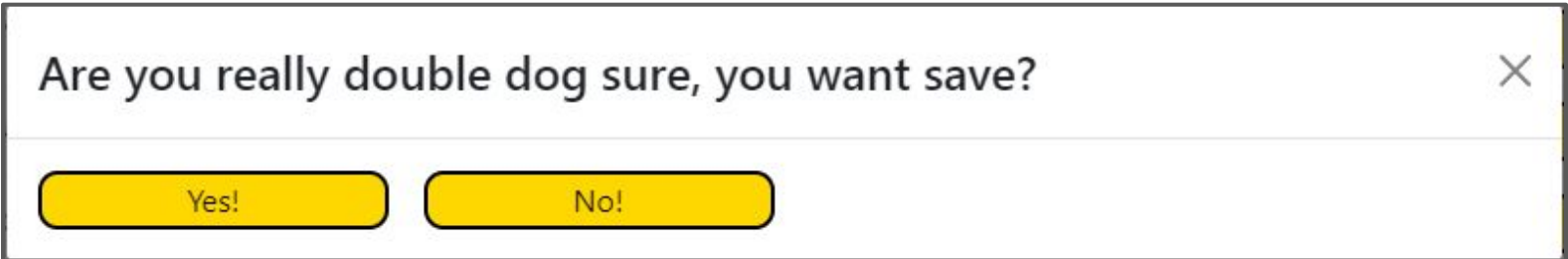
1. Button to run the parser and load in the terms
2. Button to save the terms in the output directory
3. Terms are selectable (to be removed)
4. Button to show context of the terms
5. Buttons to page through terms
6. Button to clear selected terms (in blue)
7. Button to delete selected terms

Service Walkthrough

Term Extraction Popup Windows



1. Popup window that displays context of the term (file and sentences)



1. Popup window to confirm saving (this will be used on all future save buttons)

Service Walkthrough

Category Creation

1DOCUMENTS

2TERMS

3CATEGORIES

4TAXONOMY

Category Creation

Terms

Select terms to move to category

Noun	Frequency	Weight
instance	1	1
large aircraft rule	1	1
pilot	2	2
the most technologically advanced commercial aircraft	1	1
the power	1	1
the world	1	1
they	1	1
tight maneuver	1	1
very tight turn	1	1

Edit Categories

+ Create New Category

↺ Clear Selected

💾 Save Categories

Categories

Select a category to edit

Category	Terms
Category1	<div>a 737 or other small aircraft a very large aircraft a very short approach</div>
Category2	

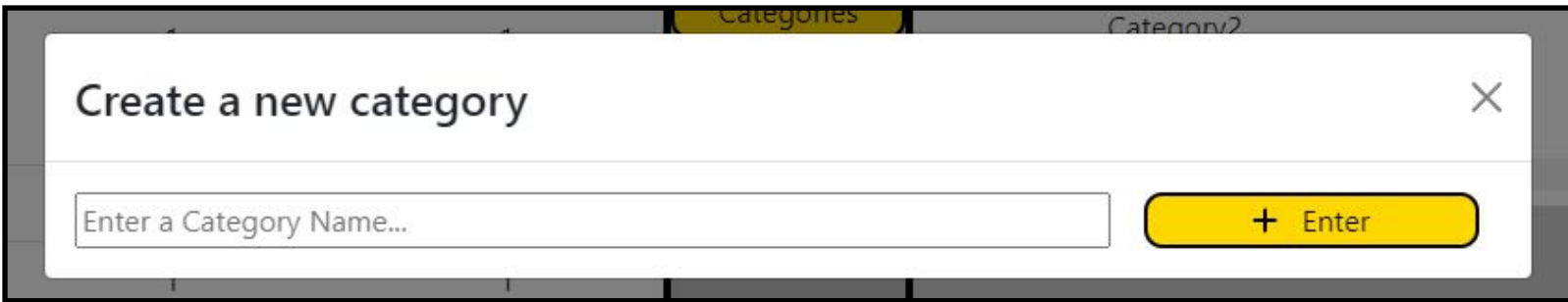
< Previous: 0 Next: >

⏪ Back Forward ⏩

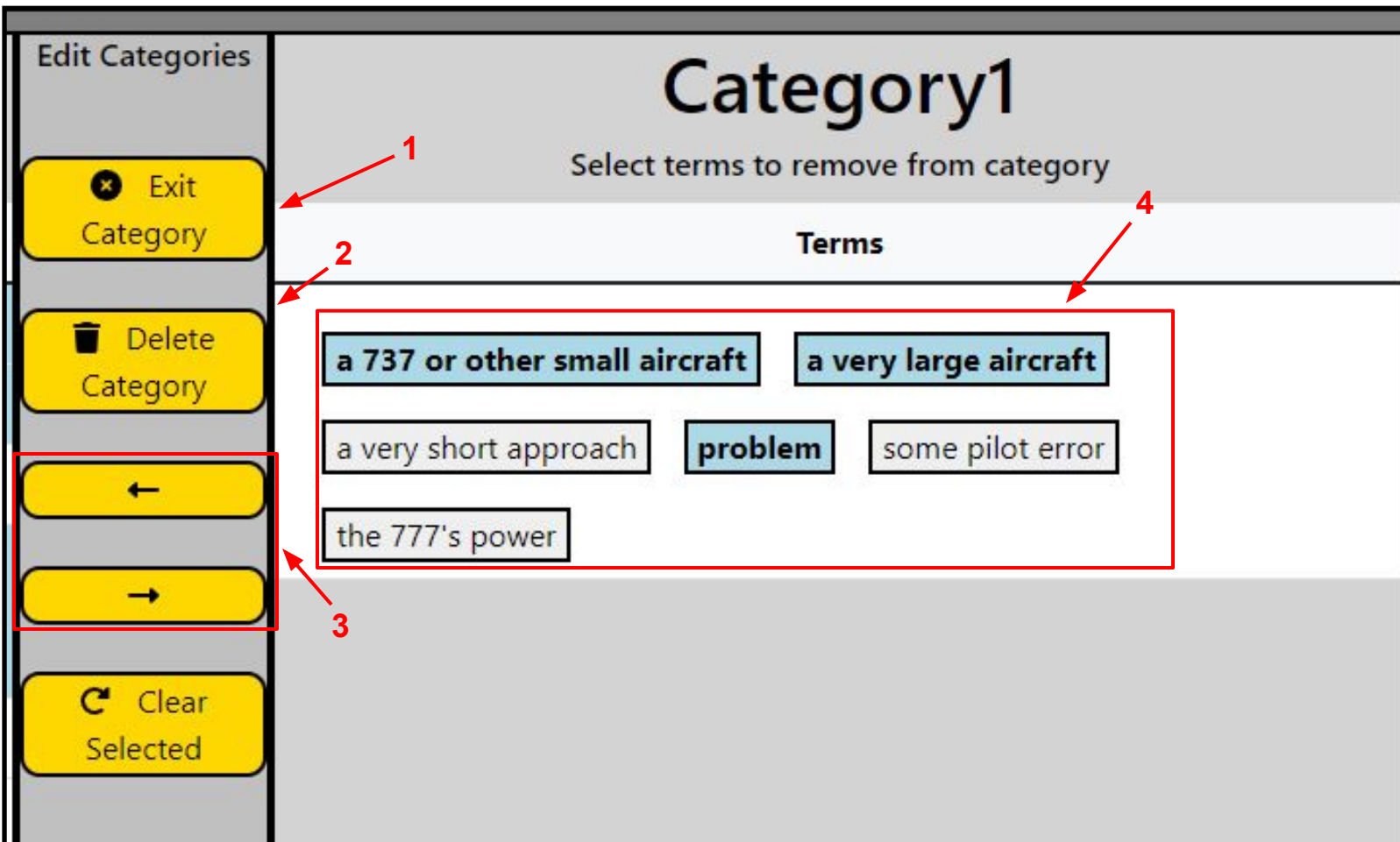
1. Terms are selectable (in blue)
2. Button to create a new category
3. Button to clear the selected terms
4. Button to save categories in output location
5. Categories are selectable
6. Terms in categories are scrollable

Service Walkthrough

Category Creation Popup Windows



1. Popup window to create a new category



1. Button to exit the category
2. Button to delete the category
3. Button to add or remove terms from the category
4. Terms in category are selectable (in blue)

Service Walkthrough

Taxonomy

1DOCUMENTS

2TERMS

3CATEGORIES

4TAXONOMY

Taxonomy

Hold ctrl to click on multiple nodes or edges

Currently Selected Edges: category2>category3
Currently Selected Nodes: 1

```
graph LR; category1 --> category2; category2 --> category3;
```

3

4

7

Create a New Relationship

+ Add Edge

See Nouns

Delete Edge

Take Screenshot

Save Relationship

Click to Edit

Relationship	Color
is-a	#000000
part-of	#ff0000

5

6

8

Back

1. Window to display currently selected categories and relationships
2. Category relationship graph can be clicked and dragged
 - a. To select multiple categories, user must hold down ctrl key
3. Button to create a new relationship type
4. Button to add an edge between two selected categories (only clickable when two categories are selected)
5. Button to see the terms in selected category (only clickable when one category is selected)
6. Button to delete the selected edges (only clickable when at least one edge is selected)
7. Button to take a screenshot and download it to your browser
8. Relationship types are selectable

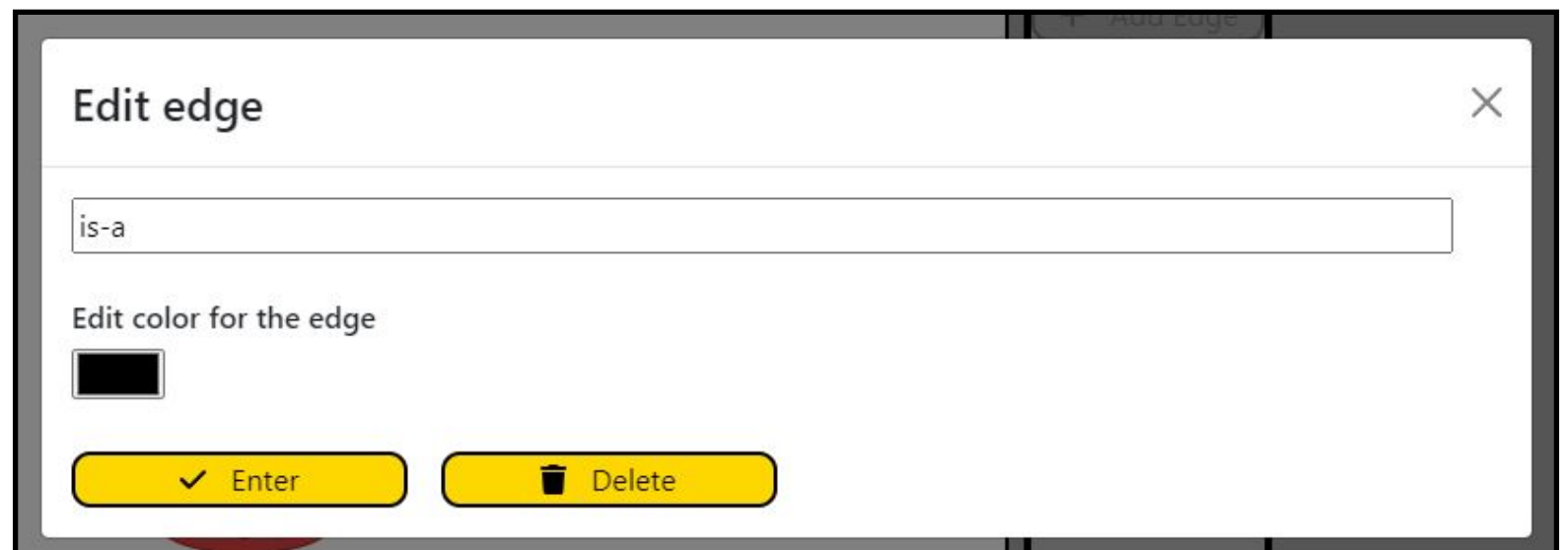
Service Walkthrough

Taxonomy Popup Windows



The screenshot shows a 'Create a new edge type' popup window. It has a title bar with a close button (X) in the top right corner. Below the title bar is a text input field with the placeholder text 'Enter a Relationship Name...'. Underneath the input field is the label 'Choose a color for the edge' followed by a green color selection button. A red arrow with the number '1' points to this button. At the bottom of the window is a yellow button with a plus icon and the text 'Create'.

1. Color select button to select a color for the relationship type



The screenshot shows an 'Edit edge' popup window. It has a title bar with a close button (X) in the top right corner. Below the title bar is a text input field containing the text 'is-a'. Underneath the input field is the label 'Edit color for the edge' followed by a black color selection button. At the bottom of the window are two yellow buttons: one with a checkmark icon and the text 'Enter', and another with a trash can icon and the text 'Delete'.

1. Popup window to edit a relationship

Service Walkthrough

Taxonomy Popup Windows

Add Relationship



helps-with ▾



+ Add

1. Drop down menu to select a relationship type

category1



Nouns

a 737 or other small aircraft

a very large aircraft

a very short approach

instance

large aircraft rule

problem

pilot

1. Popup window which display the terms in a category