Let’s Learn NLP – Student Edition 2022

User Manual

Table of Contents

[**1. Acknowledgements** 2](#_Toc97894664)

[**2. General Information** 2](#_Toc97894665)

[**3. Requirements** 2](#_Toc97894666)

[**4. Installation and Setup** 3](#_Toc97894667)

[**5. Workshop Structure** 7](#_Toc97894668)

[**6. FAQ** 8](#_Toc97894669)

[**7. Troubleshooting** 9](#_Toc97894670)

[**8. References** 9](#_Toc97894671)

# **1. Acknowledgements**

Special thanks to **Dr Chris Harding**, CEO of Lacibus Ltd. Working on a project for Lacibus has been an extremely interesting journey during which we had a chance to explore the world of Natural Language Processing and share this knowledge with younger students through this workshop.

# **2. General Information**

**Project title:** Let’s Learn NLP – Student Edition 2022

**Team Authors First Names, Surnames and Email addresses:**

Bartosz Grabek [bartosz.grabek.20@ucl.ac.uk](mailto:bartosz.grabek.20@ucl.ac.uk)

Suraj Kothari [suraj.kothari.20@ucl.ac.uk](mailto:suraj.kothari.20@ucl.ac.uk)

Minyi Lei [minyi.lei.20@ucl.ac.uk](mailto:minyi.lei.20@ucl.ac.uk)

**Summary:**

A workshop in Python Jupyter Notebooks about the fundamentals of Natural Language Processing.

**Keywords:** Python, Jupyter Notebooks, NLP, Spacy

# **3. Requirements**

**A. Technical Requirements**

* Access to a **PC** or **laptop** (**1** per student/group)
* Access to the Internet (for additional video tutorials)
* **Python** **3.9.x** installed on the computer (3.9.7 was used for development)
* **Jupyter** **notebook** package installed (**Helpful guide**: <https://jupyter.org/install>)
* Space: ~500MB
* RAM: 4GB+

**B. Other Requirements**

* We recommend the students have some basic knowledge about **programming in Python**, however, many of the exercises will involve running simple commands and we have done our best to abstract any complicated logic away.
* There will be some use of **maths** to explain how words are stored. We recommend students have some knowledge of **vectors** and basic operations on them.

# **4. Installation and Setup**

**A. Installing Python**

If you already have Python installed on your computer and its version is equal or higher than 3.9, you can skip this section. If it is below 3.9.x, we strongly recommend updating your Python by uninstalling existing version and installing a new version 3.9 or higher. To do that:

**Step 1** Go to [www.python.org](http://www.python.org)

**Step 2** Navigate to downloads and choose a Python version matching your OS

**Step 3** Download and install Python. Follow the installation instructions and remember to check the box: Add Python to PATH.



**Step 4** To check whether you have installed python correctly open a terminal window and type python –version or python3 –version (for earlier releases). You should see your version of Python printed in terminal.

**B. Installing Jupyter**

If you already have Jupyter Notebook installed on your computer, you can skip this section.

If not, open the terminal and type pip install notebook

**C. Downloading and using project code**

**Step 1** Download the tutorial package folder (.zip) and unpack it in your preferred directory

**Step 2** Open the terminal and navigate to the project folder (lets-learn-nlp-2022)

**Step 3** If you don’t have it already, install virtualenv. To do so, type:

pip install virtualenv

**Step 4** Type in the following command to create a virtual environment nlp\_env

virtualenv nlp\_env

A new folder nlp\_env will be created in project directory.

**Step 5** Activate a virtual environment

On Windows: nlp\_env/Scripts/activate

On MAC: source nlp\_env/Scripts/activate

**Step 6** Install all the necessary dependencies using this command:

pip install -r requirements.txt

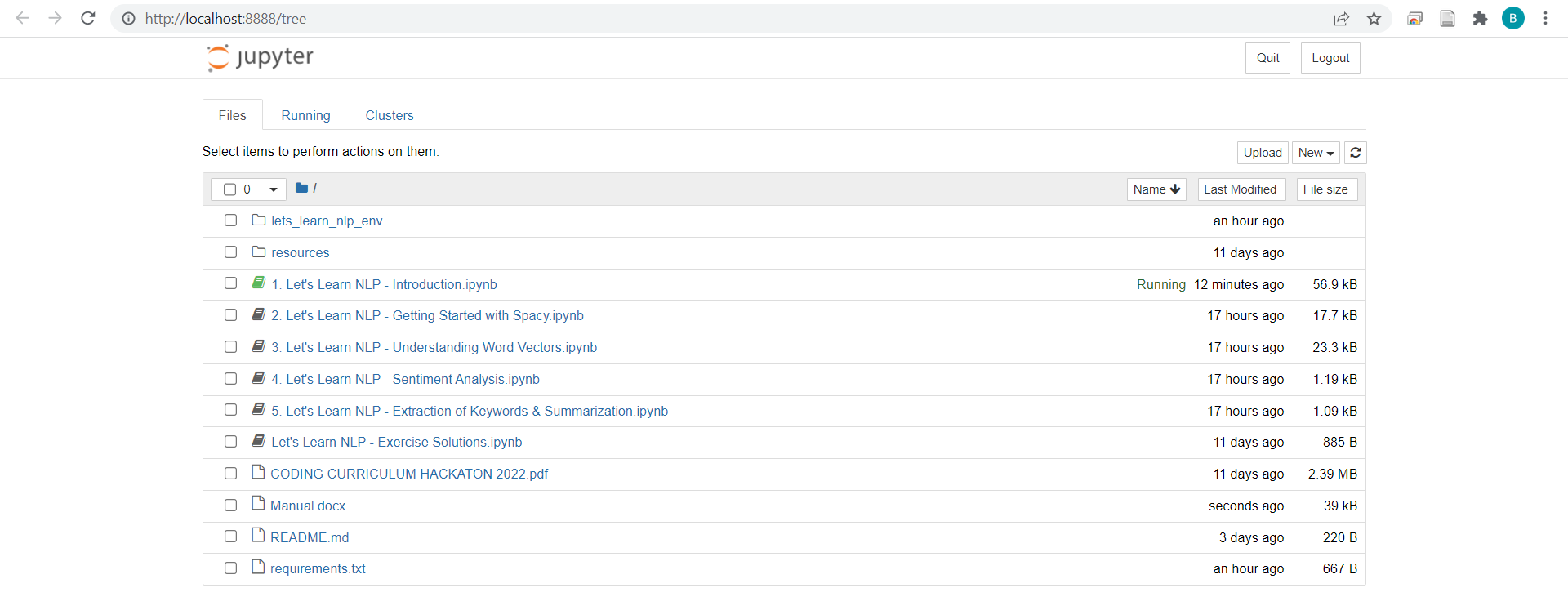
**Step 7** Install kernel. We will call it nlp\_env just like the environment.

ipython kernel install --user --name=nlp\_env

**Step 8** Run Jupyter Notebook

jupyter notebook

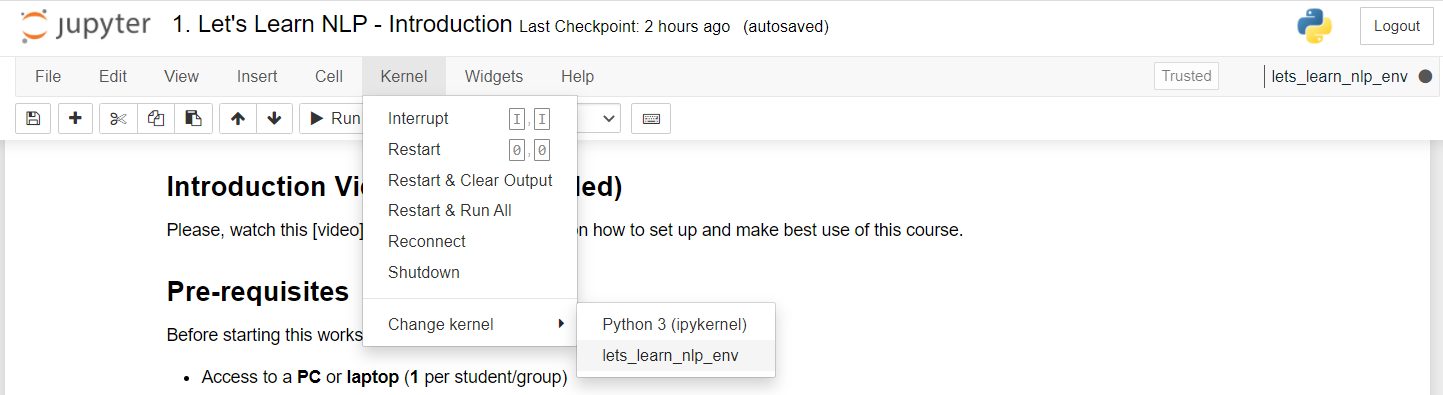
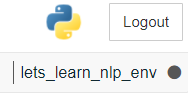
Your browser should now host the jupyter notebook and you should be able to see the folder contents inside it (see Figure 1). **Note:** Jupyter Notebook may look differently depending on your OS and default browser.



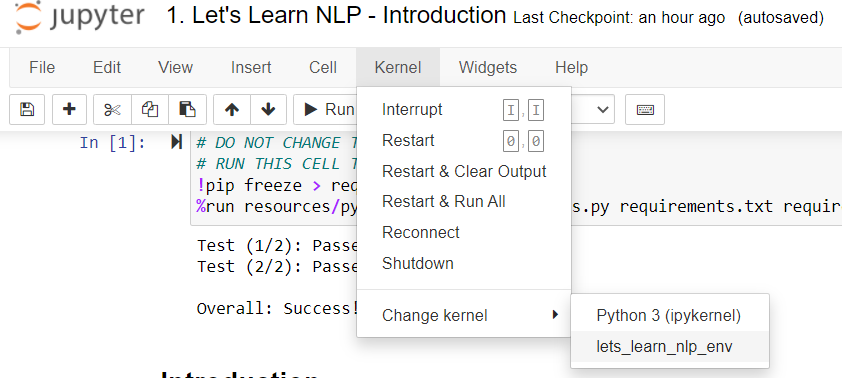
*Figure 1*

Click on the first notebook “1. Let’s Learn NLP – Introduction.ipynb” to start.

In the top right corner under the logout button you should see the name of the kernel used. Make sure its **lets\_learn\_nlp\_env** (see Figure 2). If it is different, from the top navigation bar choose “Kernel/Change kernel/lets\_learn\_nlp\_env” (see Figure 3)



*Figure 2*



*Figure 3*

When you want to close the course, just close the browser and the terminal running it.

To open it again you can simply navigate to the project folder, open the terminal, activate the environment by typing lets\_learn\_nlp\_env/Scripts/activate on Windows or source lets\_learn\_nlp\_env/Scripts/activate on MAC OS. Next, type jupyter notebook to start Jupyter Notebook.

Your setup is now completed and you are ready to use the course. Watch the “Installation and Setup Guide” video if you need any help.

**E. Uninstallation**

**Removing jupyter kernel**

Remove the newly installed kernel by opening terminal and typing

jupyter kernelspec uninstall myenv

**Deleting project folder**

You can simply delete project folder like any other folder on your computer.

**Uninstalling Python**

You can uninstall Python just like any other programme. Uninstalling Python will also remove all the libraries and its dependencies including Jupyter Notebook.

# **5. Workshop Structure**

1. **Introduction**
   1. Introduction Video
   2. Prerequisites
   3. Learning Objectives
   4. Checking Setup
   5. Introduction to NLP
   6. Getting familiar with Jupyter Notebook
   7. Video 1 (example)
   8. Quiz 1 (example)
   9. Exercise 1 (example)
2. **Getting Started with Spacy**
   1. Using spacy language Model
   2. Video 2 Spacy Pipeline
   3. Elements of Spacy Pipeline
      1. Tokenizer
      2. Using the doc object
      3. Tagger
      4. Parser
      5. Lemmatizer
      6. Named Entity Recognizer (NER)
      7. Other features (stopwords and token attributes)
   4. Quiz 2
   5. Exercise 2
3. **Understanding word vectors**
   1. Word vectors
   2. (optional) How are word vectors created?
   3. Measuring word similarity
      1. Cosine similarity / distance
      2. Exercise 3.1
      3. Exercise 3.2
      4. Exercise 3.3
   4. Calculating similarity between two words
   5. Word Vector Arithmetic
   6. Exercise 3.4
   7. Quiz 3
4. **Sentiment Analysis**
   1. What is Sentiment Analysis?
   2. How is Sentiment Analysis usually done?
   3. How will we do it?
   4. Quiz 4
   5. Exercise 4
5. **Extraction of Keywords and Summarization**
   1. What are keywords?
   2. How to extract keywords?
   3. Summarization
   4. Quiz 5
   5. Exercise 5
   6. Summary

# **6. FAQ**

**Can I download the course for free? Where can I download it?**

Yes, the course is totally free. You can either clone it using Git or download it as a .zip folder via our GitHub repository [here](https://github.com/Bartosz7/lets-learn-nlp-2022).

**How long is the course?**

The course should take you about 2-4 hours.

**Do I need Internet to use the course?**

While the course can be done fully offline, we recommend having internet access to view the tutorial videos.

**I’m a teacher. How can I use these materials?**

The course is designed to be a self-learning session for students. Teachers may assist students by answering their questions or helping them with the quizzes and exercises. Solutions to exercises are within the course materials, so students have access to them, but should only use them to assess their own solutions or in case they do not know how to solve a particular problem.

**Will you update the course?**

The course may be updated in the future. It depends on the interest and feedback we receive.

**What to do if I have a technical issue with using the course?**

If you have a GitHub account, you can report any issues on the Issues tab on our project Github repository. Before doing so, check whether you have the latest version of the course and whether you have successfully followed the Installation and Setup Guide.

If you haven’t got a GitHub account, you can use [this form](https://docs.google.com/forms/d/1g5XaNZibGYi4tAl11lUt9dXFwd-0s0H0TIne72jH5CI/edit?usp=sharing).

**How can I give feedback?**

The last chapter of the workshop includes a feedback form. The same form can be accessed via [this link](https://docs.google.com/forms/d/1f-K5esQSkh2bvvPw_5XjU-nXMOYf6X7WkLNVJ8GJh8w/edit?usp=sharing). You’re welcome to share your experience with others.

# **7. Troubleshooting**

This troubleshooting section has been created to provide guidance if any issues occur and ways to resolve them.

**Python libraries/modules are missing**

When opening a new notebook, it may have the kernel defaulted to *Python 3 – ipykernel*, however, you need to switch the kernel to *nlp\_env* to make use of the dependencies.

***nlp\_env* is not available when switching kernels**

If *nlp\_env* is not showing as an option, then the virtual environment hasn’t been added to jupyter notebook. The command to install it is: ipython kernel install --user --name=nlp\_env.

# **8. References**