



VA

U.S. Department
of Veterans Affairs

Using Python in VINCI



VA Informatics
and Computing
Infrastructure

Overview

- Basic Python Introduction
 - What is Python?
 - Why use Python?
- How to use Python in VINCI
 - Requirements
 - Installing Python
 - Adding packages
- Python use cases in VINCI
 - Projects using Python
 - Python packages developed in VINCI

Overview

- **Basic Python Introduction**
 - **What is Python?**
 - **Why use Python?**
- How to use Python in VINCI
 - Requirements
 - Installing Python
 - Adding packages
- Python use cases in VINCI
 - Projects using Python
 - Python packages developed in VINCI

What is Python?

- Multi-paradigm programming language
 - Object Oriented Programming
 - Functional Programming
 - etc.
- Simple syntax
 - Control flow determined by whitespace
 - Dynamic typing
- Open source, community funded development
- First released in 1991



```
def gcd(a, b):  
    if (b == 0):  
        return a  
    else:  
        return gcd(b, a % b)  
  
print(gcd(60, 48))  
  
>>> 12
```

Why use Python? – The Language

- Emphasizes code readability
 - Improves research transparency
 - Good for reproducibility
- Platform agnostic
 - The same code works anywhere
- Allows for quickly going from idea to working prototype

Why use Python? – The Community

- Python is a primary language for data science community
- Thousands of available packages for different algorithms or tools are maintained
 - Utilize work of others rather than creating implementations from scratch
- Community help-sites like StackOverflow and Github Issues provide quick troubleshooting

Overview

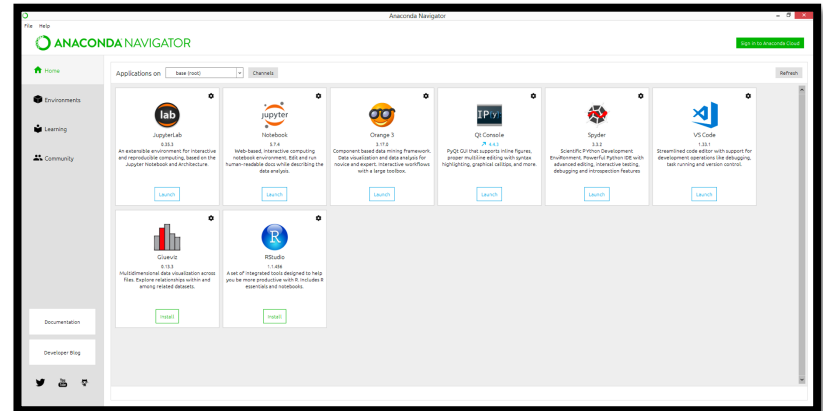
- Basic Python Introduction
 - What is Python?
 - Why use Python?
- **How to use Python in VINCI**
 - **Requirements**
 - **Installing Python**
 - **Adding packages**
- Python use cases in VINCI
 - Projects using Python
 - Python packages developed in VINCI

Requirements for using Python

- Development Workspace access (not Standard Workspace)
 - R is a good substitute for Python if only the Standard Workspace is available
 - Apply to get a Development VM at: VINCI Central -> Computing Cloud
- Access to the S: Drive with approved software
- Install Anaconda from installers available on S: drive

What is Anaconda?

- Python (and R) distribution
- Designed for scientific computing and large-scale applications
- Manages virtual environments for projects
- Handles package installation
- Provides a user interface and command line tools

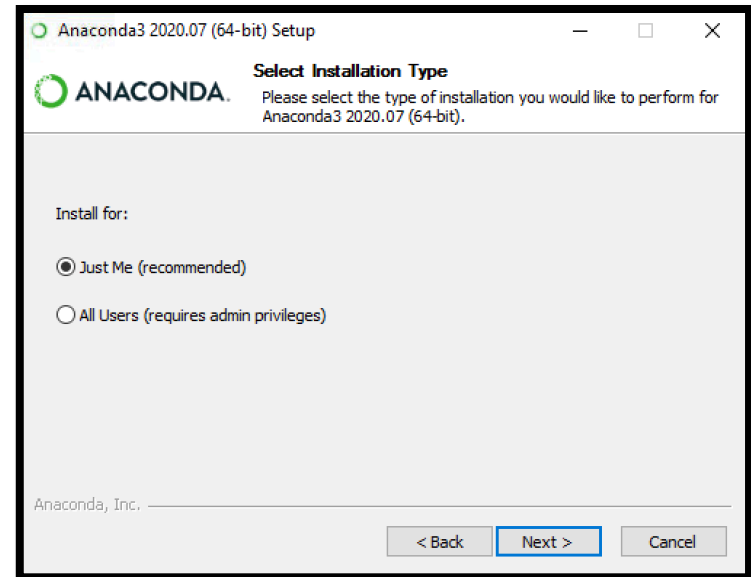


Why use Anaconda?

- Virtual environments protect your code
 - Keep packages separate
 - Allow different projects to use different versions of a package
 - Protect python from corruption
- Anaconda's Python distribution is faster than other distributions
- User interface allows users without command line experience to use Python

Installing Anaconda

- Anaconda installers available at S:\Site Licensed\Anaconda
- Anaconda can be installed **without** administrator privileges for one user at a time
- Detailed instructions available online and during installation



Packages Included in Anaconda

NumPy 

matplotlib 
Version 3.4.2

 **SciPy**

 **pandas**

 **scikit
learn**

 **seaborn**

 **Numba**

SQLAlchemy

bokeh 

Black 

 **SPYDER**

IP[y]: IPython
Interactive Computing

 **Jupyter**

Adding New Packages to VINCI

- Anaconda includes many popular general-purpose packages
- Extra installation required for specialized packages
- Installing large packages is difficult
- Requires downloading installation files or project source code outside VINCI then uploading and installing within VINCI
- Requires downloading **ALL** requirements/dependencies of the package

Adding New Packages to VINCI – Options

Manual Download

- Manually download all packages and requirements
- Slow
- Prone to user error
- No command line necessary

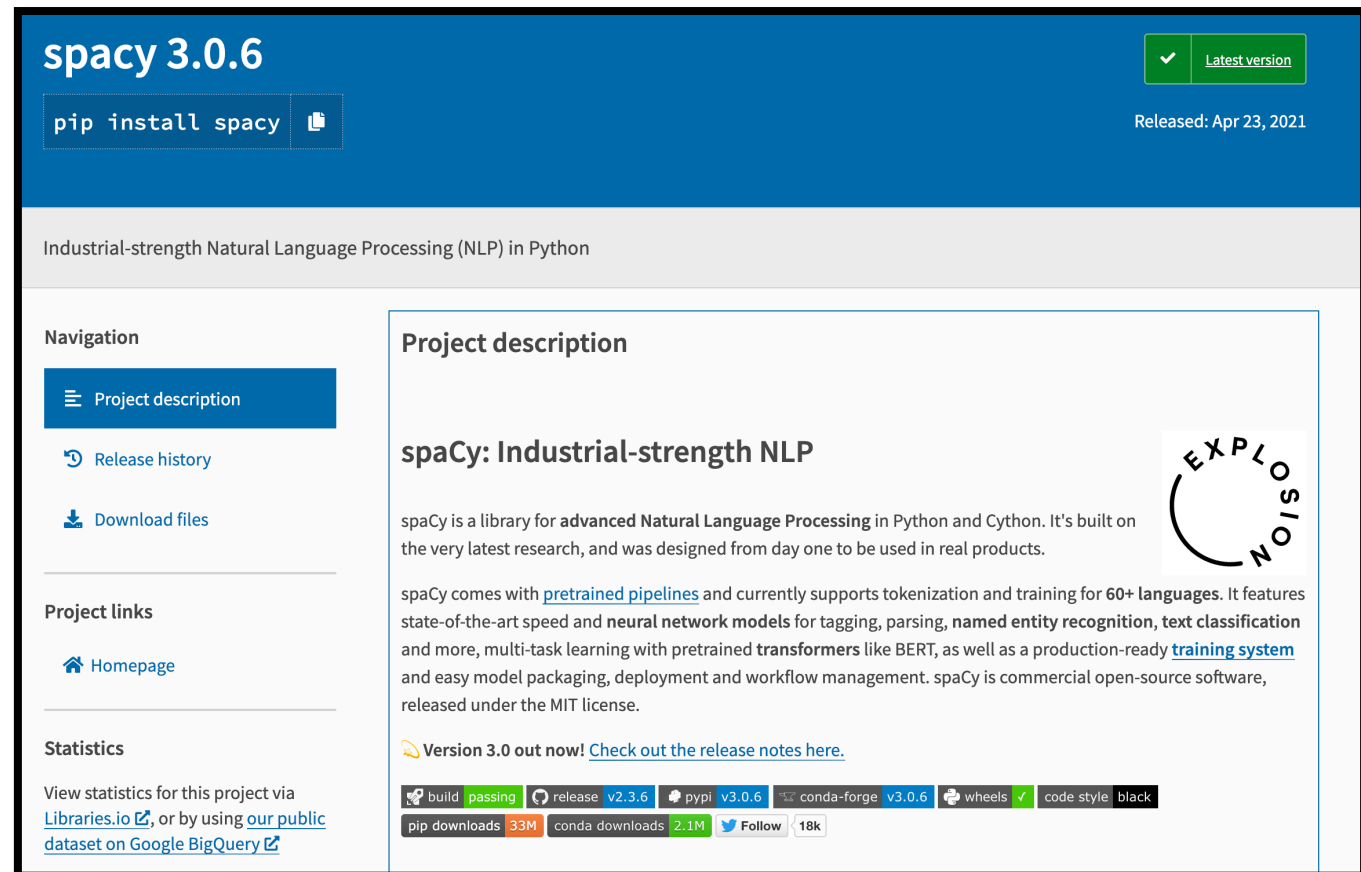
Pip Download

- Automatically download all packages and requirements
- Fast
- Sometimes downloads incompatible versions
- Using command line


Adding New Packages to VINCI – Manual Download

spaCy is a package you might want to install in VINCI.

This is the PyPI webpage for spaCy.



spacy 3.0.6 ✓ Latest version

`pip install spacy` 

Released: Apr 23, 2021

Industrial-strength Natural Language Processing (NLP) in Python

Navigation

- [Project description](#)
- [Release history](#)
- [Download files](#)

Project links

- [Homepage](#)

Statistics


View statistics for this project via [Libraries.io](#), or by using [our public dataset on Google BigQuery](#)


Project description







spaCy: Industrial-strength NLP





spaCy is a library for **advanced Natural Language Processing** in Python and Cython. It's built on the very latest research, and was designed from day one to be used in real products.

spaCy comes with [pretrained pipelines](#) and currently supports tokenization and training for 60+ languages. It features state-of-the-art speed and **neural network models** for tagging, parsing, **named entity recognition**, **text classification** and more, multi-task learning with pretrained **transformers** like BERT, as well as a production-ready [training system](#) and easy model packaging, deployment and workflow management. spaCy is commercial open-source software, released under the MIT license.



 **Version 3.0 out now!** [Check out the release notes here.](#)

Adding New Packages to VINCI – Manual Download

Click
Download Files



spacy 3.0.6

✓ Latest version

Released: Apr 23, 2021

```
pip install spacy
```

Industrial-strength Natural Language Processing (NLP) in Python

Navigation

- Project description
- Release history
- Download files

Project links

- Homepage

Statistics

View statistics for this project via [Libraries.io](#), or by using [our public dataset on Google BigQuery](#)

Project description

spaCy: Industrial-strength NLP


spaCy is a library for **advanced Natural Language Processing** in Python and Cython. It's built on the very latest research, and was designed from day one to be used in real products.

spaCy comes with [pretrained pipelines](#) and currently supports tokenization and training for **60+ languages**. It features state-of-the-art speed and **neural network models** for tagging, parsing, **named entity recognition**, **text classification** and more, multi-task learning with pretrained **transformers** like BERT, as well as a production-ready [training system](#) and easy model packaging, deployment and workflow management. spaCy is commercial open-source software, released under the MIT license.

🚀 Version 3.0 out now! [Check out the release notes here.](#)

build passing release v2.3.6 pypi v3.0.6 conda-forge v3.0.6 wheels code style black

pip downloads 33M conda downloads 2.1M Follow 18k



Adding New Packages to VINCI – Manual Download

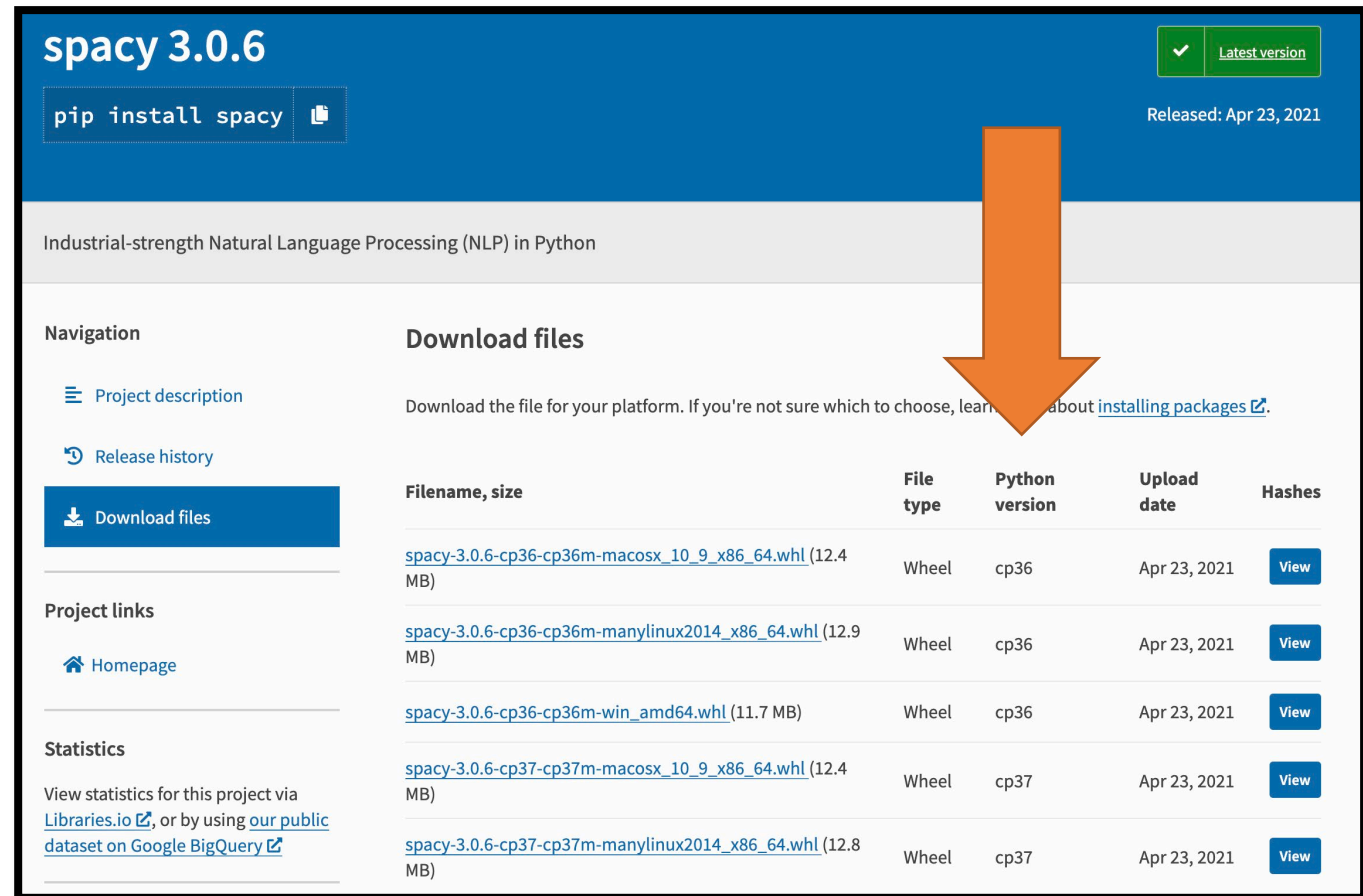
Look for the files for your Python version.

cp36 = python 3.6

cp37 = python 3.7

cp38 = python 3.8

...



spacy 3.0.6 Latest version
Released: Apr 23, 2021

`pip install spacy`

Industrial-strength Natural Language Processing (NLP) in Python

Navigation

- [Project description](#)
- [Release history](#)
- [Download files](#)**

Project links

- [Homepage](#)

Statistics

View statistics for this project via [Libraries.io](#), or by using [our public dataset on Google BigQuery](#)

Download files

Download the file for your platform. If you're not sure which to choose, learn more about [installing packages](#).

Filename, size	File type	Python version	Upload date	Hashes
spacy-3.0.6-cp36-cp36m-macosx_10_9_x86_64.whl (12.4 MB)	Wheel	cp36	Apr 23, 2021	View
spacy-3.0.6-cp36-cp36m-manylinux2014_x86_64.whl (12.9 MB)	Wheel	cp36	Apr 23, 2021	View
spacy-3.0.6-cp36-cp36m-win_amd64.whl (11.7 MB)	Wheel	cp36	Apr 23, 2021	View
spacy-3.0.6-cp37-cp37m-macosx_10_9_x86_64.whl (12.4 MB)	Wheel	cp37	Apr 23, 2021	View
spacy-3.0.6-cp37-cp37m-manylinux2014_x86_64.whl (12.8 MB)	Wheel	cp37	Apr 23, 2021	View

Adding New Packages to VINCI – Manual Download

Look at filenames to find a file compatible with your development workspace.

Any computer:

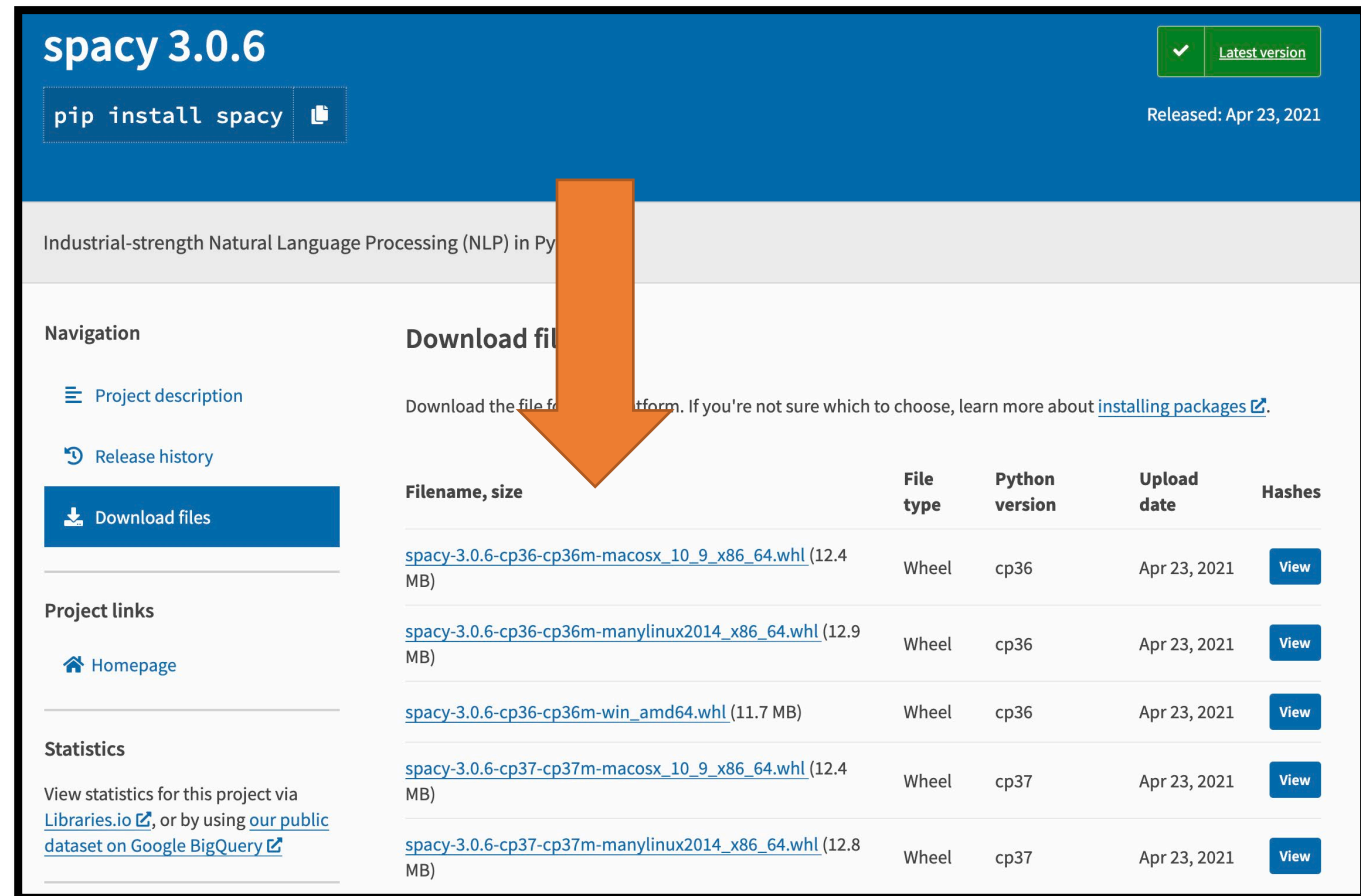
- noarch
- no platform specification

Windows:

- win_amd64

Linux:

- manylinux



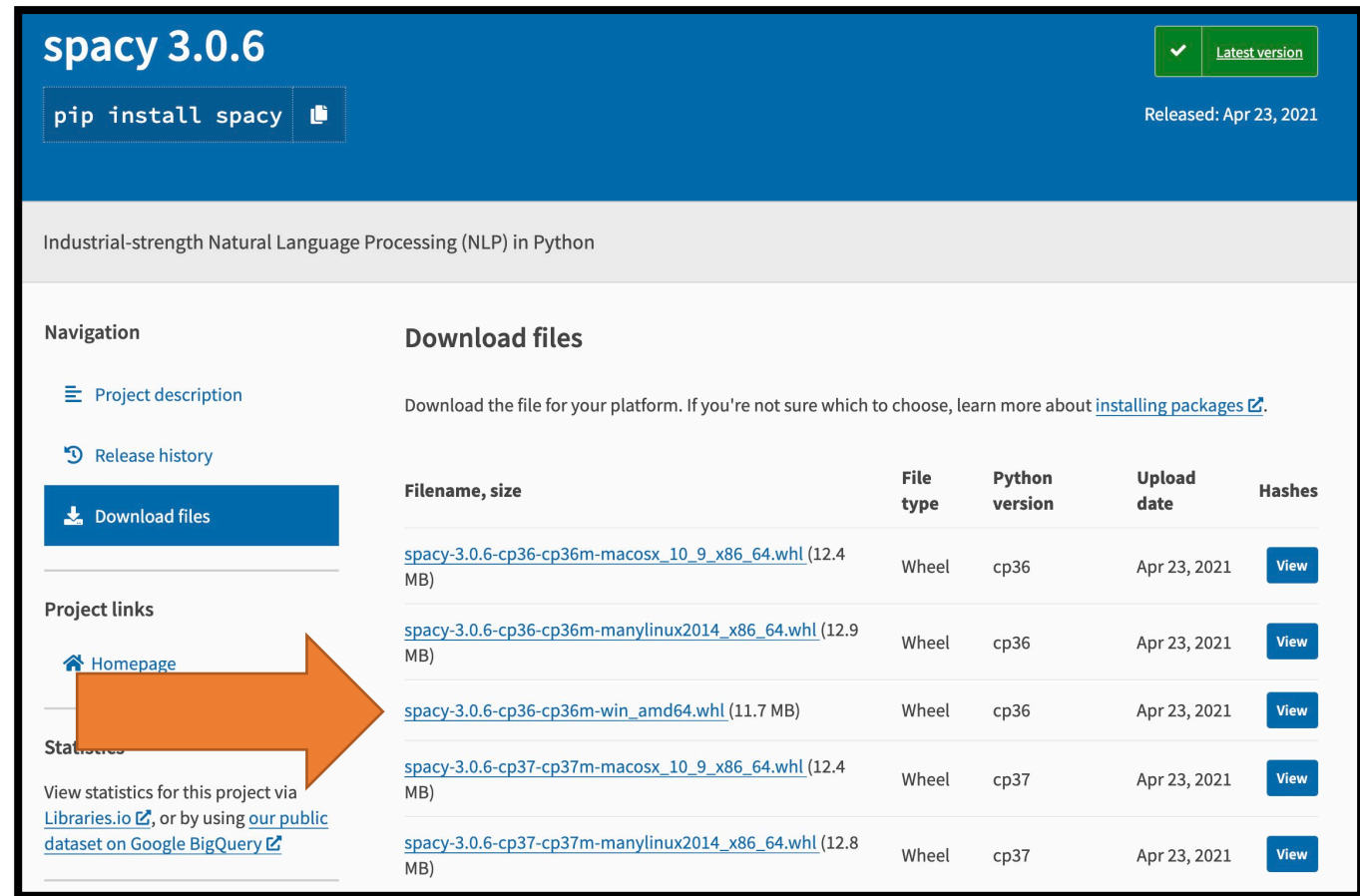
The screenshot shows the PyPI page for **spacy 3.0.6**. The page includes a navigation sidebar on the left with options like "Project description", "Release history", and "Download files". The main content area is titled "Download files" and contains a table of available wheel files. A large orange arrow points to the "Filename, size" column of the table.

Filename, size	File type	Python version	Upload date	Hashes
spacy-3.0.6-cp36-cp36m-macosx_10_9_x86_64.whl (12.4 MB)	Wheel	cp36	Apr 23, 2021	View
spacy-3.0.6-cp36-cp36m-manylinux2014_x86_64.whl (12.9 MB)	Wheel	cp36	Apr 23, 2021	View
spacy-3.0.6-cp36-cp36m-win_amd64.whl (11.7 MB)	Wheel	cp36	Apr 23, 2021	View
spacy-3.0.6-cp37-cp37m-macosx_10_9_x86_64.whl (12.4 MB)	Wheel	cp37	Apr 23, 2021	View
spacy-3.0.6-cp37-cp37m-manylinux2014_x86_64.whl (12.8 MB)	Wheel	cp37	Apr 23, 2021	View


Adding New Packages to VINCI – Manual Download

Download the file for your computer **AND** python version.

This file is for a Windows workspace using Python 3.6



spacy 3.0.6 ✓ Latest version

`pip install spacy` 

Released: Apr 23, 2021

Industrial-strength Natural Language Processing (NLP) in Python

Navigation

- [Project description](#)
- [Release history](#)
- [Download files](#)

Project links

- [Homepage](#)

Statistics

View statistics for this project via [Libraries.io](#), or by using [our public dataset on Google BigQuery](#)

Download files

Download the file for your platform. If you're not sure which to choose, learn more about [installing packages](#).

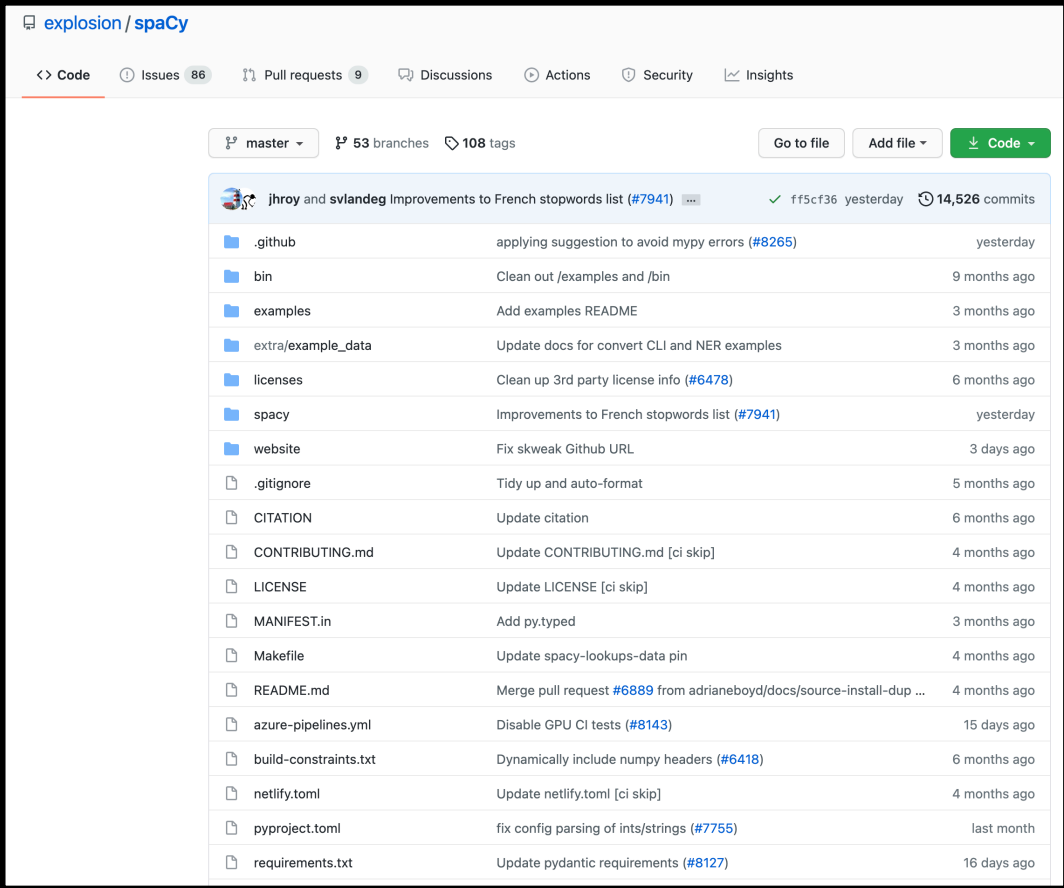
Filename, size	File type	Python version	Upload date	Hashes
spacy-3.0.6-cp36-cp36m-macosx_10_9_x86_64.whl (12.4 MB)	Wheel	cp36	Apr 23, 2021	View
spacy-3.0.6-cp36-cp36m-manylinux2014_x86_64.whl (12.9 MB)	Wheel	cp36	Apr 23, 2021	View
spacy-3.0.6-cp36-cp36m-win_amd64.whl (11.7 MB)	Wheel	cp36	Apr 23, 2021	View
spacy-3.0.6-cp37-cp37m-macosx_10_9_x86_64.whl (12.4 MB)	Wheel	cp37	Apr 23, 2021	View
spacy-3.0.6-cp37-cp37m-manylinux2014_x86_64.whl (12.8 MB)	Wheel	cp37	Apr 23, 2021	View

Adding New Packages to VINCI – Manual Download

Now to get dependencies.

The easiest way to look at this is to open the Github page for the same project.

You can also open the file you just downloaded (but save a copy for later).

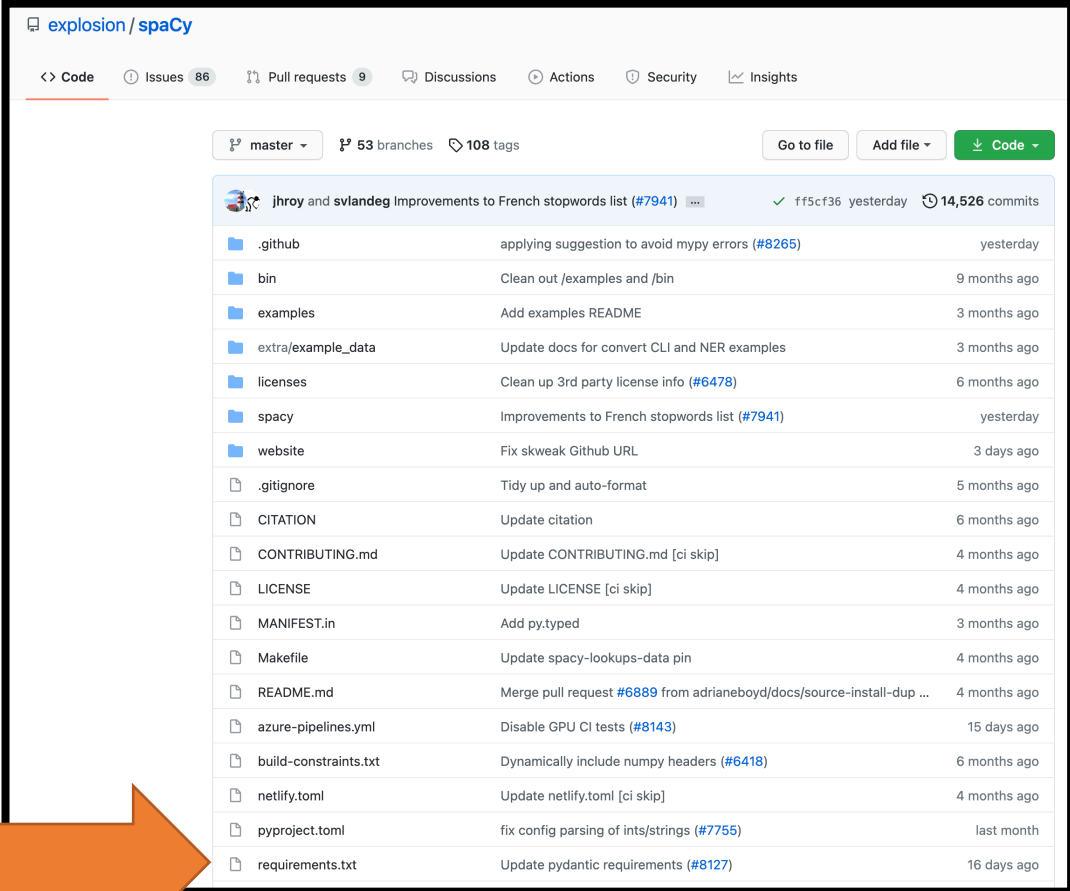


The screenshot shows the GitHub repository page for `explosion/spaCy`. The repository is on the `master` branch, with 53 branches and 108 tags. The commit history is displayed, showing a list of files and their corresponding commit messages and dates.

File	Commit Message	Commit Date
<code>.github</code>	applying suggestion to avoid mypy errors (#8265)	yesterday
<code>bin</code>	Clean out /examples and /bin	9 months ago
<code>examples</code>	Add examples README	3 months ago
<code>extra/example_data</code>	Update docs for convert CLI and NER examples	3 months ago
<code>licenses</code>	Clean up 3rd party license info (#6478)	6 months ago
<code>spacy</code>	Improvements to French stopwords list (#7941)	yesterday
<code>website</code>	Fix skweak Github URL	3 days ago
<code>.gitignore</code>	Tidy up and auto-format	5 months ago
<code>CITATION</code>	Update citation	6 months ago
<code>CONTRIBUTING.md</code>	Update CONTRIBUTING.md [ci skip]	4 months ago
<code>LICENSE</code>	Update LICENSE [ci skip]	4 months ago
<code>MANIFEST.in</code>	Add py.typed	3 months ago
<code>Makefile</code>	Update spacy-lookups-data pin	4 months ago
<code>README.md</code>	Merge pull request #6889 from adrianeboyd/docs/source-install-dup ...	4 months ago
<code>azure-pipelines.yml</code>	Disable GPU CI tests (#8143)	15 days ago
<code>build-constraints.txt</code>	Dynamically include numpy headers (#6418)	6 months ago
<code>netlify.toml</code>	Update netlify.toml [ci skip]	4 months ago
<code>pyproject.toml</code>	fix config parsing of ints/strings (#7755)	last month
<code>requirements.txt</code>	Update pydantic requirements (#8127)	16 days ago

Adding New Packages to VINCI – Manual Download

Open requirements.txt



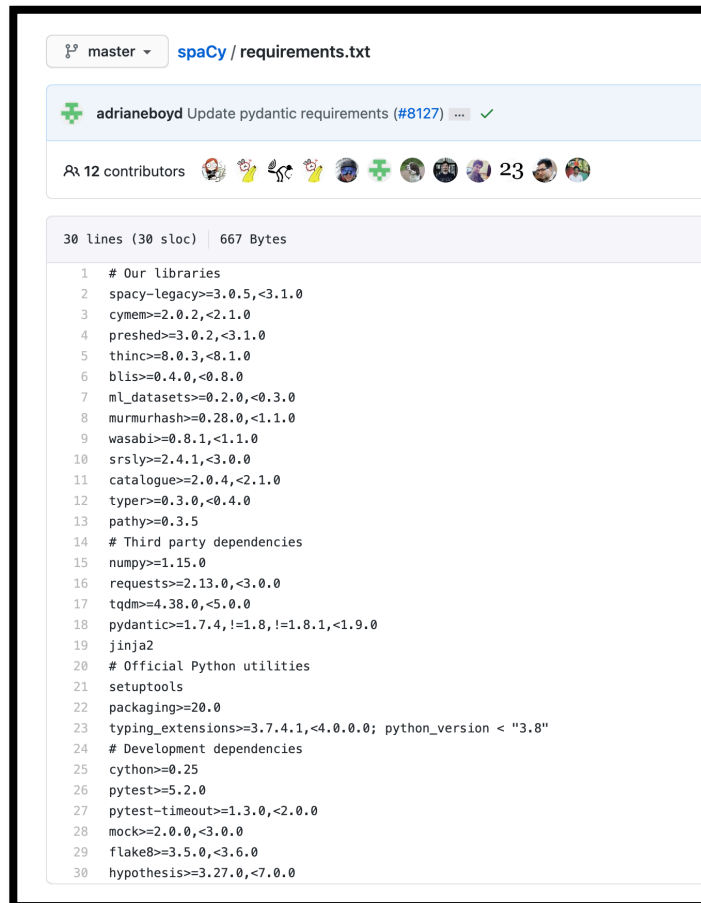
The screenshot shows the GitHub repository page for `explosion/spaCy`. The repository is on the `master` branch, with 53 branches and 108 tags. The commit history is displayed, showing a list of files and folders with their respective commit messages and dates. The `requirements.txt` file is highlighted at the bottom of the list, and an orange arrow points to it from the left.

File/Folder	Commit Message	Commit Date
<code>.github</code>	applying suggestion to avoid mypy errors (#8265)	yesterday
<code>bin</code>	Clean out /examples and /bin	9 months ago
<code>examples</code>	Add examples README	3 months ago
<code>extra/example_data</code>	Update docs for convert CLI and NER examples	3 months ago
<code>licenses</code>	Clean up 3rd party license info (#6478)	6 months ago
<code>spacy</code>	Improvements to French stopwords list (#7941)	yesterday
<code>website</code>	Fix skweak Github URL	3 days ago
<code>.gitignore</code>	Tidy up and auto-format	5 months ago
<code>CITATION</code>	Update citation	6 months ago
<code>CONTRIBUTING.md</code>	Update CONTRIBUTING.md [ci skip]	4 months ago
<code>LICENSE</code>	Update LICENSE [ci skip]	4 months ago
<code>MANIFEST.in</code>	Add py.typed	3 months ago
<code>Makefile</code>	Update spacy-lookups-data pin	4 months ago
<code>README.md</code>	Merge pull request #6889 from adrianeboyd/docs/source-install-dup ...	4 months ago
<code>azure-pipelines.yml</code>	Disable GPU CI tests (#8143)	15 days ago
<code>build-constraints.txt</code>	Dynamically include numpy headers (#6418)	6 months ago
<code>netlify.toml</code>	Update netlify.toml [ci skip]	4 months ago
<code>pyproject.toml</code>	fix config parsing of ints/strings (#7755)	last month
<code>requirements.txt</code>	Update pydantic requirements (#8127)	16 days ago

Adding New Packages to VINCI – Manual Download

Download all listed packages **AND** their requirements using the same steps.

This process can be long for packages like spaCy with lots of features, developers, and users.



The screenshot shows a GitHub pull request for the file 'requirements.txt' in the 'spaCy' repository, specifically on the 'master' branch. The pull request is titled 'Update pydantic requirements (#8127)' and is authored by 'adrianeboyd'. It has 12 contributors and 23 reviews. The file content is as follows:

```
1 # Our libraries
2 spacy-legacy>=3.0.5,<3.1.0
3 cymem>=2.0.2,<2.1.0
4 preshed>=3.0.2,<3.1.0
5 thinc>=8.0.3,<8.1.0
6 blis>=0.4.0,<0.8.0
7 ml_datasets>=0.2.0,<0.3.0
8 murmurhash>=0.28.0,<1.1.0
9 wasabi>=0.8.1,<1.1.0
10 srsly>=2.4.1,<3.0.0
11 catalogue>=2.0.4,<2.1.0
12 typer>=0.3.0,<0.4.0
13 pathy>=0.3.5
14 # Third party dependencies
15 numpy=1.15.0
16 requests>=2.13.0,<3.0.0
17 tqdm>=4.38.0,<5.0.0
18 pydantic>=1.7.4,!=1.8,!=1.8.1,<1.9.0
19 jinja2
20 # Official Python utilities
21 setuptools
22 packaging>=20.0
23 typing_extensions>=3.7.4.1,<4.0.0.0; python_version < "3.8"
24 # Development dependencies
25 cython>=0.25
26 pytest>=5.2.0
27 pytest-timeout>=1.3.0,<2.0.0
28 mock>=2.0.0,<3.0.0
29 flake8>=3.5.0,<3.6.0
30 hypothesis>=3.27.0,<7.0.0
```

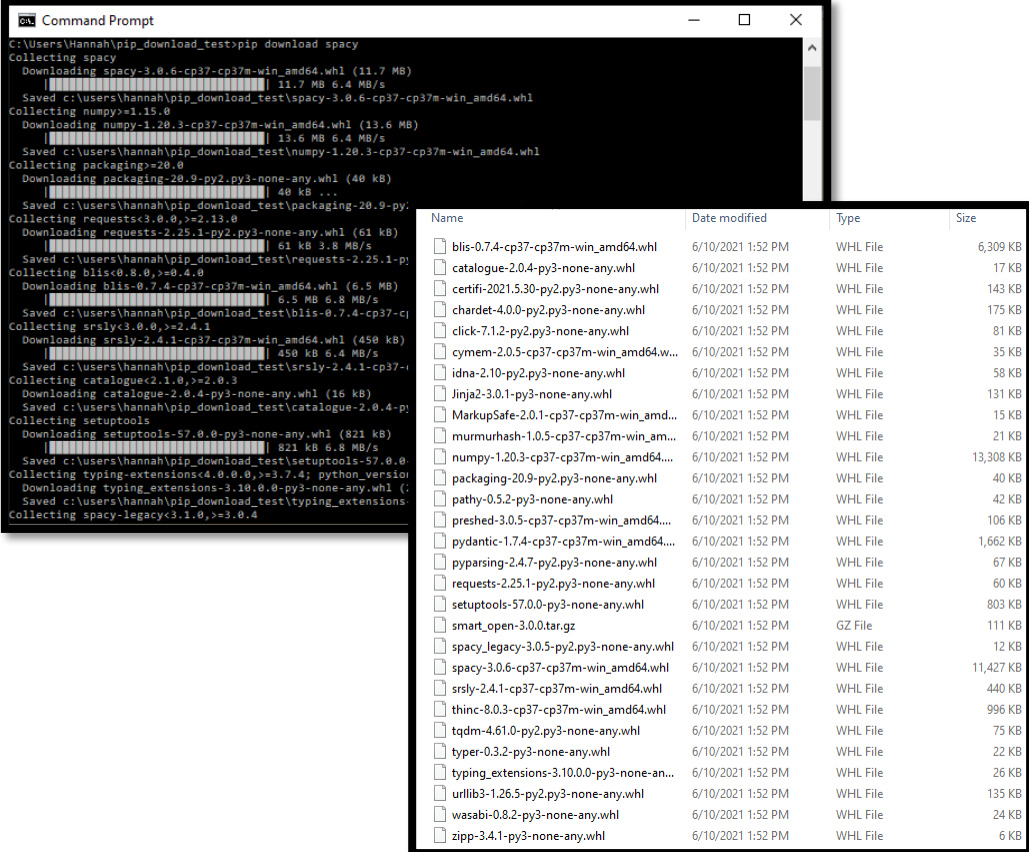
Adding New Packages to VINCI – Pip Download

Must use command line, no UI available for pip

`pip download <your_package>`

`pip download spacy`

Downloads the package and all requirements into the current location.



```
Command Prompt
C:\Users\Hannah\pip_download_test>pip download spacy
Collecting spacy
  Downloading spacy-3.0.6-cp37-cp37m-win_amd64.whl (11.7 MB)
    |#####| 11.7 MB 6.4 MB/s
  Saved c:\users\hannah\pip_download_test\spacy-3.0.6-cp37-cp37m-win_amd64.whl
Collecting numpy>=1.15.0
  Downloading numpy-1.20.3-cp37-cp37m-win_amd64.whl (13.6 MB)
    |#####| 13.6 MB 6.4 MB/s
  Saved c:\users\hannah\pip_download_test\numpy-1.20.3-cp37-cp37m-win_amd64.whl
Collecting packaging>=20.0
  Downloading packaging-20.9-py2.py3-none-any.whl (40 kB)
    |#####| 40 kB ...
  Saved c:\users\hannah\pip_download_test\packaging-20.9-py-
Collecting requests<3.0.0,>=2.13.0
  Downloading requests-2.25.1-py2.py3-none-any.whl (61 kB)
    |#####| 61 kB 3.8 MB/s
  Saved c:\users\hannah\pip_download_test\requests-2.25.1-p
Collecting blis<0.8.0,>=0.4.0
  Downloading blis-0.7.4-cp37-cp37m-win_amd64.whl (6.5 MB)
    |#####| 6.5 MB 6.8 MB/s
  Saved c:\users\hannah\pip_download_test\blis-0.7.4-cp37-c
Collecting srsly<3.0.0,>=2.4.1
  Downloading srsly-2.4.1-cp37-cp37m-win_amd64.whl (450 kB)
    |#####| 450 kB 6.4 MB/s
  Saved c:\users\hannah\pip_download_test\srsly-2.4.1-cp37-
Collecting catalogue<2.1.0,>=2.0.3
  Downloading catalogue-2.0.4-py3-none-any.whl (16 kB)
  Saved c:\users\hannah\pip_download_test\catalogue-2.0.4-p
Collecting setuptools
  Downloading setuptools-57.0.0-py3-none-any.whl (821 kB)
    |#####| 821 kB 6.8 MB/s
  Saved c:\users\hannah\pip_download_test\setuptools-57.0.0
Collecting typing-extensions<4.0.0.0,>=3.7.4; python_version
  Downloading typing_extensions-3.10.0.0-py3-none-any.whl (
  Saved c:\users\hannah\pip_download_test\typing_extensions
Collecting spacy-legacy<3.1.0,>=3.0.4
```

Name	Date modified	Type	Size
blis-0.7.4-cp37-cp37m-win_amd64.whl	6/10/2021 1:52 PM	WHL File	6,309 KB
catalogue-2.0.4-py3-none-any.whl	6/10/2021 1:52 PM	WHL File	17 KB
certifi-2021.5.30-py2.py3-none-any.whl	6/10/2021 1:52 PM	WHL File	143 KB
chardet-4.0.0-py2.py3-none-any.whl	6/10/2021 1:52 PM	WHL File	175 KB
click-7.1.2-py2.py3-none-any.whl	6/10/2021 1:52 PM	WHL File	81 KB
cymem-2.0.5-cp37-cp37m-win_amd64.w...	6/10/2021 1:52 PM	WHL File	35 KB
idna-2.10-py2.py3-none-any.whl	6/10/2021 1:52 PM	WHL File	58 KB
Jinja2-3.0.1-py3-none-any.whl	6/10/2021 1:52 PM	WHL File	131 KB
MarkupSafe-2.0.1-cp37-cp37m-win_amd...	6/10/2021 1:52 PM	WHL File	15 KB
murmurhash-1.0.5-cp37-cp37m-win_am...	6/10/2021 1:52 PM	WHL File	21 KB
numpy-1.20.3-cp37-cp37m-win_amd64...	6/10/2021 1:52 PM	WHL File	13,308 KB
packaging-20.9-py2.py3-none-any.whl	6/10/2021 1:52 PM	WHL File	40 KB
pathy-0.5.2-py3-none-any.whl	6/10/2021 1:52 PM	WHL File	42 KB
reshed-3.0.5-cp37-cp37m-win_amd64...	6/10/2021 1:52 PM	WHL File	106 KB
pydantic-1.7.4-cp37-cp37m-win_amd64...	6/10/2021 1:52 PM	WHL File	1,662 KB
pyarsing-2.4.7-py2.py3-none-any.whl	6/10/2021 1:52 PM	WHL File	67 KB
requests-2.25.1-py2.py3-none-any.whl	6/10/2021 1:52 PM	WHL File	60 KB
setuptools-57.0.0-py3-none-any.whl	6/10/2021 1:52 PM	WHL File	803 KB
smart_open-3.0.0.tar.gz	6/10/2021 1:52 PM	GZ File	111 KB
spacy_legacy-3.0.5-py2.py3-none-any.whl	6/10/2021 1:52 PM	WHL File	12 KB
spacy-3.0.6-cp37-cp37m-win_amd64.whl	6/10/2021 1:52 PM	WHL File	11,427 KB
srsly-2.4.1-cp37-cp37m-win_amd64.whl	6/10/2021 1:52 PM	WHL File	440 KB
thinc-8.0.3-cp37-cp37m-win_amd64.whl	6/10/2021 1:52 PM	WHL File	996 KB
tqdm-4.61.0-py2.py3-none-any.whl	6/10/2021 1:52 PM	WHL File	75 KB
typer-0.3.2-py3-none-any.whl	6/10/2021 1:52 PM	WHL File	22 KB
typing_extensions-3.10.0.0-py3-none-an...	6/10/2021 1:52 PM	WHL File	26 KB
urllib3-1.26.5-py2.py3-none-any.whl	6/10/2021 1:52 PM	WHL File	135 KB
wasabi-0.8.2-py3-none-any.whl	6/10/2021 1:52 PM	WHL File	24 KB
zipp-3.4.1-py3-none-any.whl	6/10/2021 1:52 PM	WHL File	6 KB

Adding New Packages to VINCI – Pip Download

Convenient **ONLY** when you have matching platforms.

Will have errors or download incorrect versions when downloading on a different operating system.

This is an error that appears when downloading windows spaCy on a Mac



```
pip_download_test — -zsh — 80x24
Using cached https://files.pythonhosted.org/packages/e7/f8/62520edb641dde8ba57
f7ba9aa82f3c8e6567b8b8aacb690615c9800d156/srslly-2.4.1-cp37-cp37m-win_amd64.whl
Saved ./srslly-2.4.1-cp37-cp37m-win_amd64.whl
Collecting tqdm<5.0.0,>=4.38.0 (from spacy)
Using cached https://files.pythonhosted.org/packages/42/d7/f357d98e9b50346bcb6
095fe3ad205d8db3174eb5edb03edfe7c4099576d/tqdm-4.61.0-py2.py3-none-any.whl
Saved ./tqdm-4.61.0-py2.py3-none-any.whl
Collecting typing-extensions<4.0.0.0,>=3.7.4; python_version < "3.8" (from spacy
)
Using cached https://files.pythonhosted.org/packages/2e/35/6c4fff5ab443b57116c
b1aad46421fb719bed2825664e8fe77d66d99bcbc/typing_extensions-3.10.0.0-py3-none-an
y.whl
Saved ./typing_extensions-3.10.0.0-py3-none-any.whl
Collecting zipp>=0.5; python_version < "3.8" (from catalogue<2.1.0,>=2.0.3->spac
y)
Using cached https://files.pythonhosted.org/packages/0f/8c/715c54e9e34c0c4820f
616a913a7de3337d0cd79074dd1bed4dd840f16ae/zipp-3.4.1-py3-none-any.whl
Saved ./zipp-3.4.1-py3-none-any.whl
Collecting smart-open<4.0.0,>=2.2.0 (from pathy>=0.3.5->spacy)
ERROR: Could not find a version that satisfies the requirement smart-open<4.0.
0,>=2.2.0 (from pathy>=0.3.5->spacy) (from versions: 4.1.2, 5.0.0, 5.1.0)
ERROR: No matching distribution found for smart-open<4.0.0,>=2.2.0 (from pathy=>
0.3.5->spacy)
(base) u0280859@IM-EPI365 pip_download_test %
```


Adding New Packages to VINCI

Once all packages are downloaded

Navigate to VINCI Central

Click Quick Links



The screenshot shows the VINCI Central dashboard. At the top left is the VINCI Central logo. The main header reads "Welcome to the VA Informatics and Computing Infrastructure" and includes a "Keyword Search" box with a "Search" button. On the top right, contact information for Hannah Eyre is provided: VINCI Email: VINCI@VA.gov, Concierge Desk: 1-801-872-3324, Computer or General Help: Your IT, and VINCI Online Help.

A left-hand navigation menu lists various links: Launch Workspace, New To VINCI, My VINCI Dashboard, VINCI University, VINCI Tube, VINCIpedia/FAQ, User Guides, VINCI Services, Computing Cloud, Data Sources, Applications, Policies and Forms, Let's Collaborate VAI, Quick Links, and Google Notes. Below this menu, a notice states: "VINCI Monthly Server Updates on Sunday, June 13th from 10AM to 4PM ET".

The main content area features a "New To VINCI" banner with the text "Learn more about VINCI, VINCI Studies and the virtual environment." Below the banner is a grid of quick links, each with an icon and a brief description:

- VINCI University**: Here you learn more
- VINCI Tube**: Videos on all things VINCI
- VINCIpedia/FAQ**: VINCI's forum for research & info
- User Guides**: Let us guide you through VINCI
- VINCI Services**: Support and concierge services
- Computing Cloud**: Computers, software, and storage
- Data Sources**: Learn about our available data
- Applications**: Applications available in VINCI

On the right side of the dashboard, there is a "My VINCI Dashboard" section with a user icon and the text "All your VINCI resources in one place".

Adding New Packages to VINCI

Then click Upload Files



Quick Links

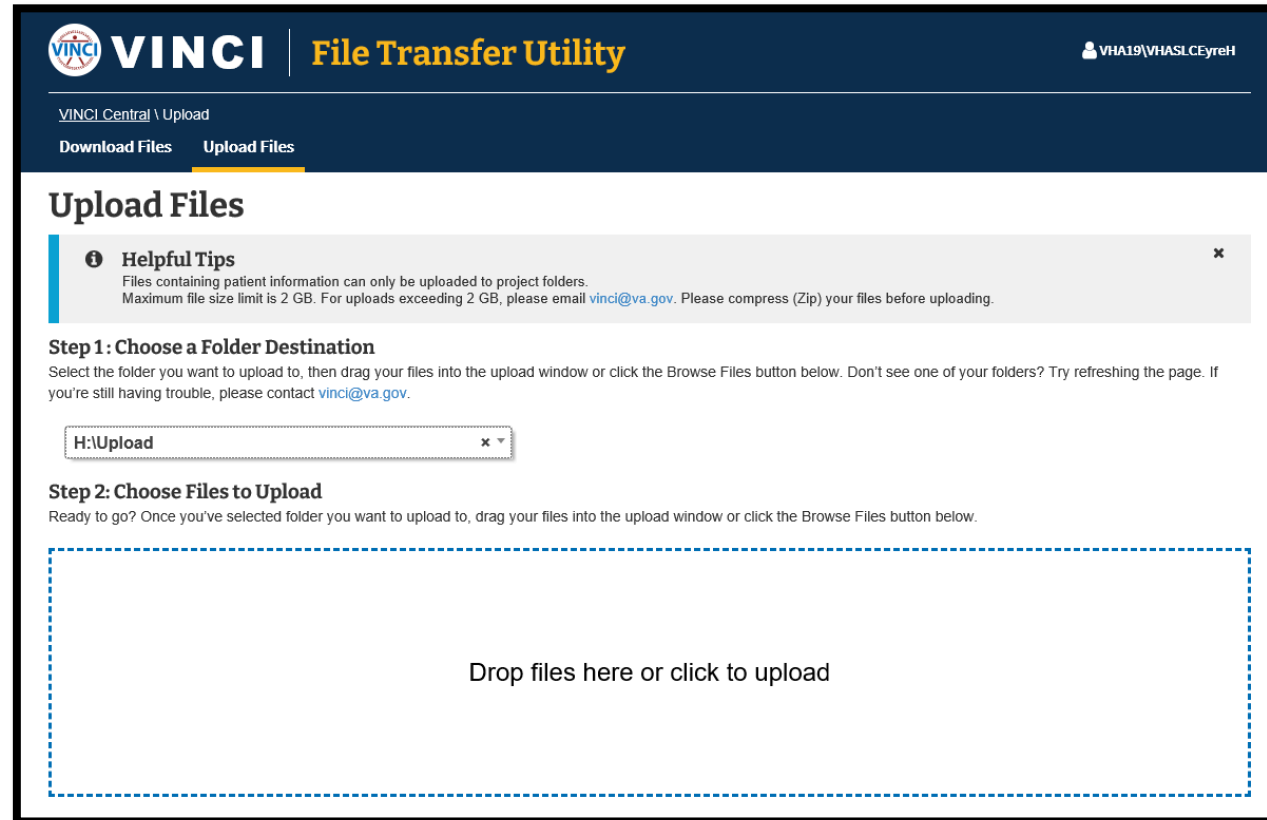
- [BaseCamp](#)
- [CDW Support Wiki](#)
- [CDW/BISL web site](#)
- [Centralized Interactive Phenomics Resource \(CIPHER\)](#)
- [COVID-19 Shared Data Resource SharePoint site](#)
- [DART](#)
- [Dim Data Viewer](#)
- [Download Files](#)
- [HERC web site](#)
- [HSR&D web site](#)
- [HSRData-L Listserv](#)
- [Meta Data Viewer](#)
- [Request Management](#)
- [SAS Grid Guides](#)
- [Upload Files](#)
- [VHA Data Portal web site](#)
- [VHA Support Service Center \(VSSC\)](#)
- [VINCI Git Services](#)
- [VINCI Linux Wiki](#)
- [VINCI Operations](#)
- [VIRec web site](#)

Adding New Packages to VINCI

Upload all downloaded packages with VINCI Upload tool.

Can be uploaded one at a time or zipped together

Pip install after uploading



The screenshot displays the VINCI File Transfer Utility interface. At the top, the VINCI logo and 'File Transfer Utility' are visible, along with a user profile 'VHA19\WHASLCEyreH'. The breadcrumb path is 'VINCI Central \ Upload', and the 'Upload Files' tab is selected. The main heading is 'Upload Files'. A 'Helpful Tips' section states: 'Files containing patient information can only be uploaded to project folders. Maximum file size limit is 2 GB. For uploads exceeding 2 GB, please email vinci@va.gov. Please compress (Zip) your files before uploading.' Below this, 'Step 1: Choose a Folder Destination' instructs users to select a folder and drag files into the upload window. A text box shows 'H:\Upload'. 'Step 2: Choose Files to Upload' instructs users to drag files into the upload window or click the 'Browse Files' button. A large dashed blue box contains the text 'Drop files here or click to upload'.

COMING SOON – VINCI Anaconda Environment

- A pre-made environment will be available on the S drive made by VINCI NLP team
- Usable on any Windows development workspace after installing Anaconda
- Unzip environment in your development workspace and use as-is
- Includes many packages beyond basic Anaconda packages
- Updates and support will be limited, but it can help skip upload process

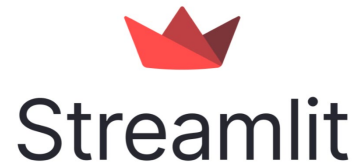
COMING SOON – VINCI Anaconda Environment

Some of the additional packages included in the VINCI Anaconda Environment



XGBoost

spaCy



mlflow



How has Python already been used?

- Completed research and operational projects within VINCI
- Created datasets in the COVID-19 Shared Data Resource
- Communicated research results with interactive displays
- Developed open-source applications shared outside the VA

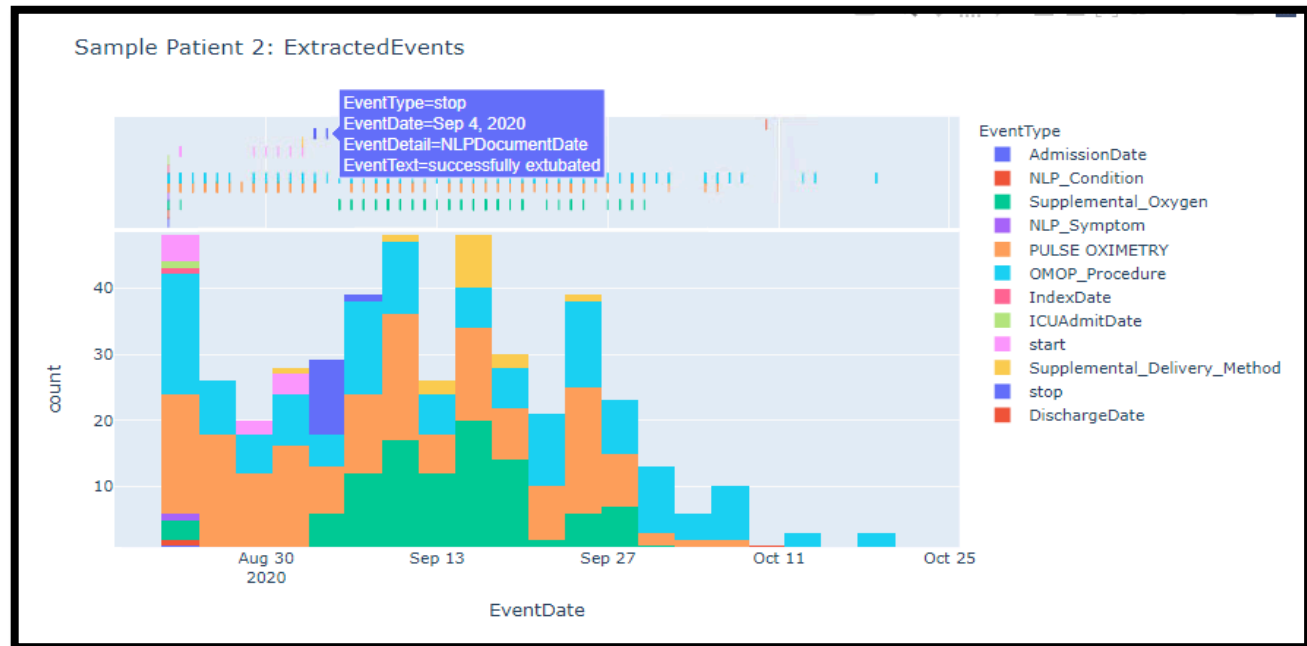
Data Visualization – Analysis

Visual representation can help debug or troubleshoot analysis code



Clear diagrams help understand the data intuitively

Can easily combine many data sources into one image

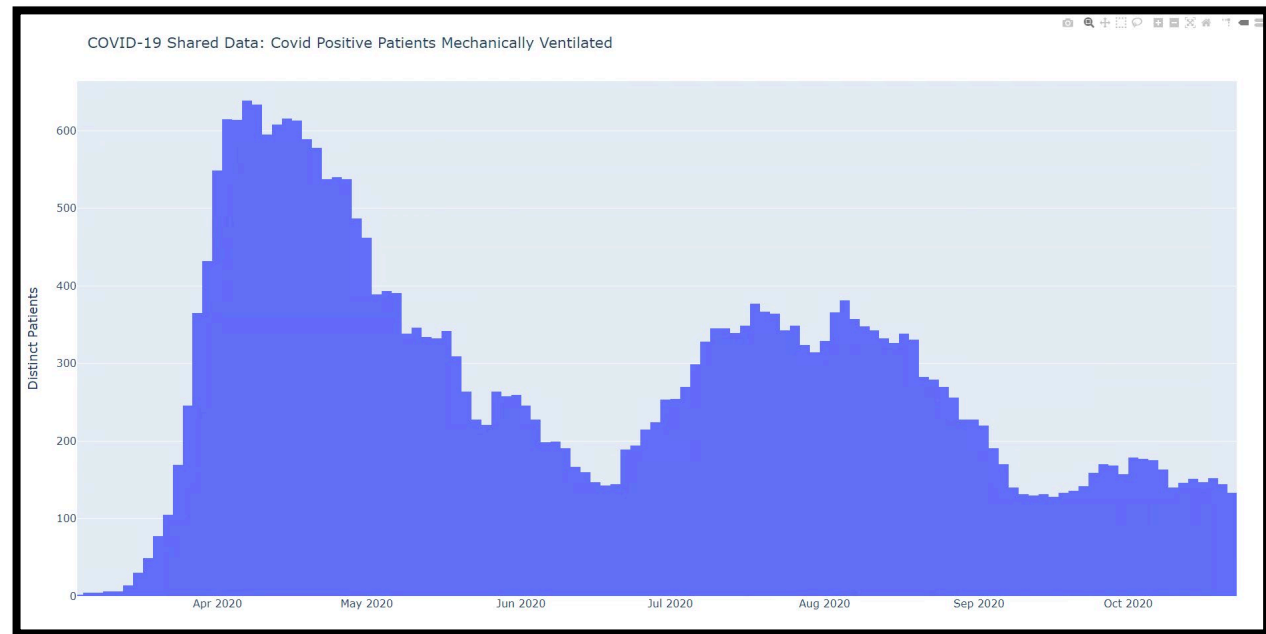


Data Visualization - Communication

Aggregated or cleaned up data used to communicate research results

Useful for dashboards that are understood by non-analyst team members

Results can be interactive

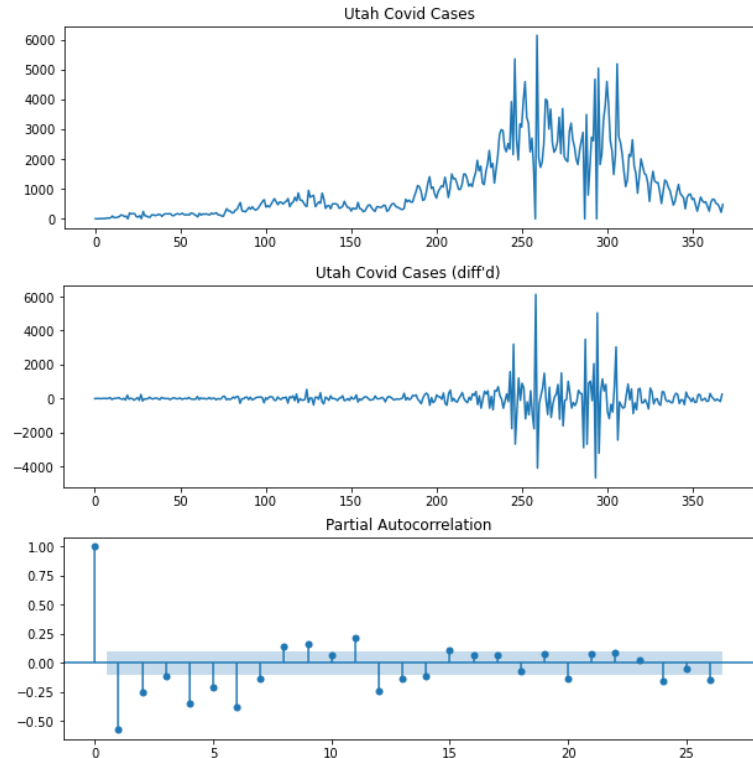


Data Science

Use mathematical and statistical techniques to understand VA data

Mathematical libraries allow powerful modeling techniques to be used

Optimized libraries allow for high performance on extremely large quantities of data



Natural Language Processing – Word Embeddings

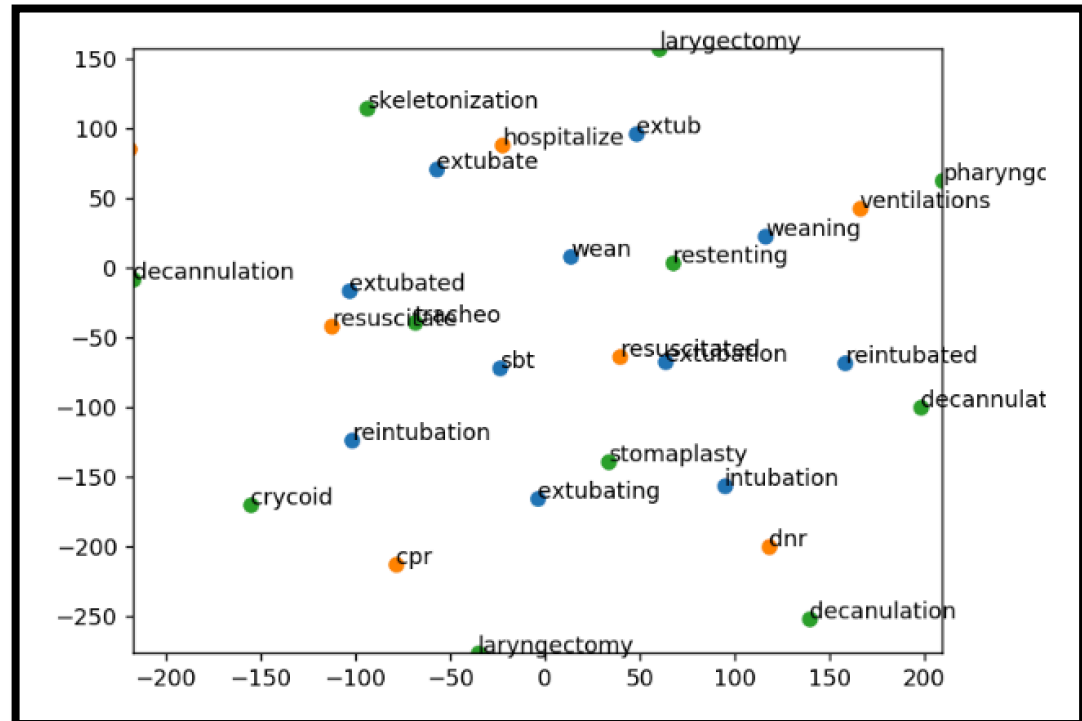
Train word embeddings on raw VA clinical notes

Use up to hundreds of millions of notes



Use for text similarity and vocab expansion

Next word prediction in a sequence



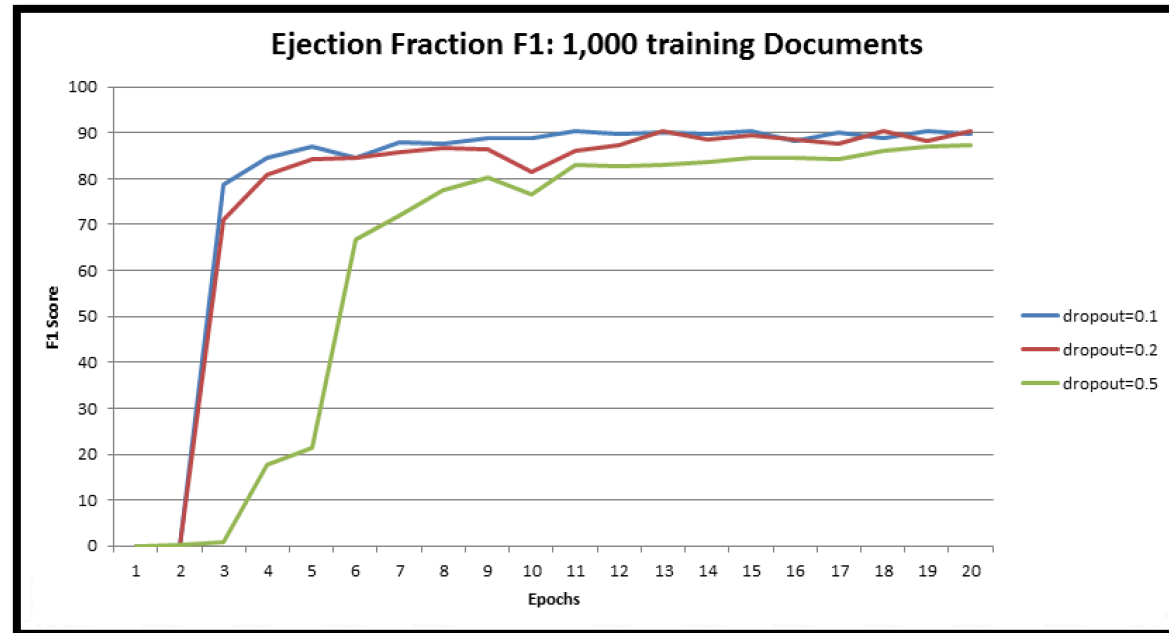
Natural Language Processing – Deep Learning

Train convolutional neural network models on a variety of tasks

Evaluate against a large test set

Create deployment-ready models to apply on future data

spaCy



Natural Language Processing - medspaCy

- medspaCy is an NLP toolkit based on spaCy
 - spaCy is the most popular general-purpose NLP package in Python
- Designed for performing common clinical NLP tasks in Python
- Compatible with other spaCy components
- Easy to incorporate other Python libraries for tasks like deep learning, visualization, and parallel deployment

Natural Language Processing - medspaCy

- medspaCy has several ways to identify clinical concepts
- Custom rule-based matching
- UMLS mapping with QuickUMLS
- Allows for use of spaCy's CNN models or models from any other source package

Family History:

Mother **FAMILY** with **stroke PROBLEM** at age 82. no early deaths.

2 daughters- healthy

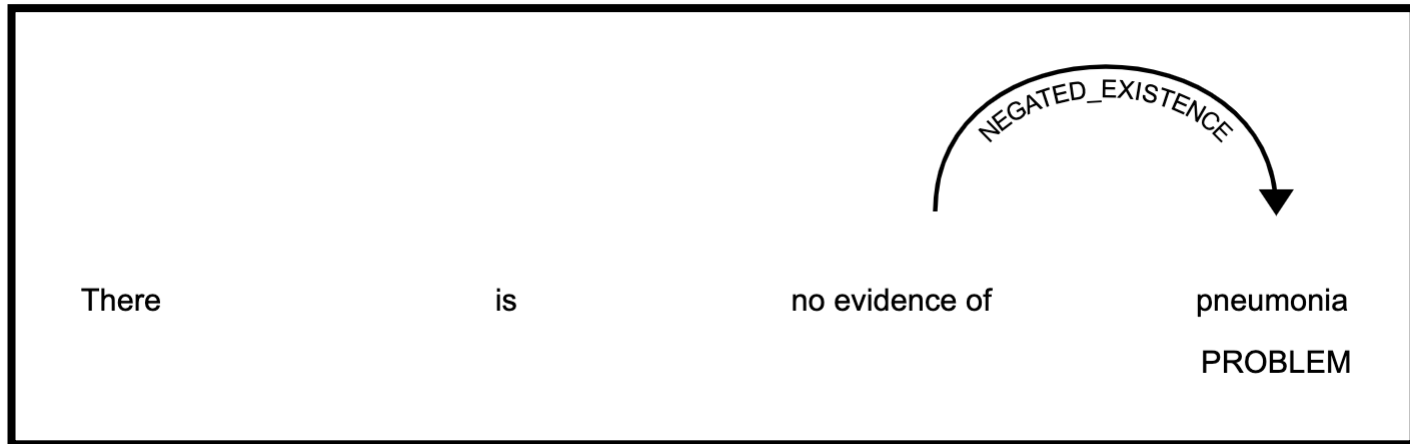
Natural Language Processing - medspaCy

- medspaCy offers an algorithm for clinical section detection
- Rule-based section patterns
- Large set of default rules available, but complete customization possible
- Can create section hierarchies for sections and subsections

Family History: << FAMILY_HISTORY >>
Mother FAMILY with stroke PROBLEM at age 82. no early deaths.
2 daughters- healthy

Natural Language Processing - medspaCy

- medspaCy implements the ConText algorithm
- Allows for detection of common attributes such as negation, temporality, or experienter
- Can be customized for special attributes as needed



Natural Language Processing - medspaCy

- medspaCy has many other utilities
- Pre- and Post-Processing rules
- Output to Pandas dataframes and databases
- Template pipelines for deploying with all reading/writing from SQL

Natural Language Processing - medspaCy

- medspacy is used to help create data in the COVID-19 Shared Data Resource
- Identification of COVID-19 positive VA patients for VA Biosurveillance
- Is incorporated into an R library called Clinspacy
- Used for classes at University of Melbourne
- Bio-NLP parser at University of Washington
- Available on GitHub and PyPI

Questions?

- Additional Trainings and Cyberseminars:
 - VINCI Central
 - Related HSR&D Cyberseminars
 - Introduction to NLP
 - Introduction to Python
- Other resources
 - Anaconda documentation online for using conda environments
 - StackOverflow/Github Issues for troubleshooting individual packages
 - Tutorials available through individual packages through Jupyter notebooks and youtube videos
- Helpdesk: VINCI@va.gov