

# Heinz College SkillScraper

---

Do you have only 24 units left of electives and want to make sure your choices are aligned with employer demands? Let the Heinz College SkillScraper help you build the skills you need to land your dream job! The Heinz College SkillScraper is a one-stop shop for:

- Researching current employment trends and salary stats (**Bureau of Labor Statistics**)
- Learning about hot skills in today's marketplace (**Payscale.com**)
- Aggregating job board listings (**Indeed.com**)
- Choosing the most relevant courses to these demands, based on course descriptions and keywords (**Heinz Course Catalog**)

## Group 7 Members

---

- Jonathan Dyer, [jondyer@cmu.edu](mailto:jondyer@cmu.edu)
- Devraj Kori, [dkori@andrew.cmu.edu](mailto:dkori@andrew.cmu.edu)
- Muriel Pokol, [mpokol@andrew.cmu.edu](mailto:mpokol@andrew.cmu.edu)
- Brian Rhindress, [brhindre@andrew.cmu.edu](mailto:brhindre@andrew.cmu.edu)
- Matthew Samach, [msamach@andrew.cmu.edu](mailto:msamach@andrew.cmu.edu)

## Installation Instructions

---

Community modules used in the Heinz SkillScraper include:

- SpaCy: <https://spacy.io/usage/>
- Pandas: <https://anaconda.org/conda-forge/pandas>
- Tkinter: <https://docs.python.org/3/library/tkinter.html>
- BeautifulSoup: <https://pypi.org/project/beautifulsoup4/>

Installation for the above modules is straightforward:

1. Ensure you have the latest version of Python (3.7) installed on your machine, preferably via Anaconda.

2. Open your Anaconda prompt (or any command prompt if not using Anaconda) and activate whatever environment you plan to use.

3. Run the following command (replacing `conda` with `pip` if not using Anaconda) and select `[y]` or press enter when prompted:

```
conda install pandas beautifulsoup4 spacy lxml
```

4. Now run the following command **as Administrator** (sudo on \*nix, run Anaconda prompt "as Administrator" on Windows):

```
python -m spacy download en
```

That's it! You're ready to run the SkillScraper application.

## Run Instructions

After installing necessary modules, simply run the **SkillScraper.py** program to launch the GUI. The integrated master controller within the GUI will call other necessary modules. Select a job from the drop-down menu and wait for the information!



SkillScraper

## Screenshots and Video

Demo video:

<https://youtu.be/ve7diLk9Or0>

### Dream job?

Database Administrators	Submit
-------------------------	--------

### Skills Needed:

Skill		
Job_Matches		
SQL	26	
Administration	20	
Administration	20	
Recovery	19	

### Courses:

['95-799: Linux & Open Source']
['94-828: Survey Design']
['93-836: International Perspectives on Managing Arts and Heritage']
['94-876: Computational Forensics']

### Job Information:

For Database Administrators
Total Employment: 113,690
Annual Salary 25 Percentile: 63,850
Annual Median Salary: 87,020
Annual Salary 75 Percentile: 112,700






### Current Postings in Pittsburgh,PA

0	SQL Database Administrator, General
1	Operations
2	Benefits Administrator II, Zimm
3	Procurement Contract Administra
4	Database Administrator

Exit

## Sub-modules in this project:

---

-  **SkillScraper.py**: The Main Controller. This script creates the Heinz SkillScraper GUI, allows for user input, and draws on data from each of the other modules.
-  **bls\_data\_frame.py**: This script generates a data frame, list of BLS-tracked jobs, and a list of salary and employment figures for a selected job.
-  **payscale\_scraper.py**: This scrapes payscale.com for an up-to-date list of payscale.com skill listings. Not used in production version.
-  **match\_indeed\_to\_skill.py**: Scrapes the top 5-pages from a selected indeed.com job search. Cross-compares and counts the number of job skill occurrences, based on payscale.com skill lists.
-  **heinz\_scraper.py**: This script scrapes the current Heinz course catalog, parses course descriptions, and compares keywords against key skill-words identified in the indeed.com scrape. It then returns a list of all courses matching the job-skill pair.

## License

---

Carnegie Mellon University  
Heinz College of Information Systems and Public Policy  
Heinz SkillScraper ©