CS Project Overview - LinkedIn Scraper Tool

Andrew Jerome, Soham Desai, Arik Gitman, Darsh Patel, Dylan Shah

Professor Kahanda

May 5, 2025

Rutgers Alumni LinkedIn Scraper

An easy-to-use tool that helps find Rutgers alumni on LinkedIn by name and graduation year.

What is it?

This tool finds and matches the best LinkedIn profile for a Rutgers graduate. Just enter a name and graduation year, and it searches the web for the best match. It helps explore career paths, work locations, and gives a rough income estimate. You can search up to 5,000 people in one session.

Manual Search Support

You can search for one person at a time for better accuracy. This is helpful when someone has a common name or you're looking for specific results.

Search Engine Options

The tool mainly uses Google search through SerpAPI. If SerpAPI doesn't work (because of limits or errors), it switches to Bing. SerpAPI usually gives better results, especially for job titles and location info.

Fuzzy Matching and Cosine Similarity

To improve accuracy, the tool uses fuzzy matching and cosine similarity.
- Fuzzy matching helps catch name variations, like 'Jon' vs. 'Jonathan'.
- Cosine similarity checks how similar two descriptions are, even if they use different words.

Job and Income Estimation

The tool guesses someone's income and work location based on their job title. If the income seems too high (like over $500K), it marks it as 'Unknown' so the data stays realistic.

Simple, Clear Interface

The design is easy to use and clearly organized. You can see search progress in real time, stop or

restart searches, and download the results.

Exporting Results

All search results can be downloaded as a CSV file, making it easy to analyze or share later.

What Works Without SerpAPI

Even if SerpAPI is not available, these features still work:

- Bing-based search

- Fuzzy matching and cosine similarity

- Graduation year filtering

- Full user interface and CSV download options

What Needs SerpAPI

Some advanced features require SerpAPI:

- More accurate location data

- Income estimates from job titles

- Structured job information from Google search results

Sliders and Default Settings

The tool includes sliders to control how strict the match should be. A higher value means the match must be closer. The default setting (0.8) is a good balance-accurate without being too strict.

Project Status

The app is stable, fully functional, and available at:

https://linkedin-extractor.onrender.com

We plan to improve accuracy even more and support larger searches.