



9/18/2024

# Cybersecurity Jobs Data Extraction From Indeed API

CyNatGlobal Cybersecurity Jobs Data Extraction



Jermaine Lucas - Omari Watson  
PROTEXXA

# Table of Contents

## Contents

Project Breakdown: Cybersecurity Jobs Data Extraction from Indeed API ...	2
Key Features .....	2
Code Analysis.....	3
Resources/Research.....	5
Conclusion.....	5

## Project Breakdown:

### Cybersecurity Jobs Data Extraction from Indeed API

This project automates the extraction of cybersecurity job postings from the Indeed API using Python. It fetches data related to job opportunities in the cybersecurity sector and processes it into a structured CSV format for easy analysis. The script can be customized to target specific locations, roles, and experience levels. This offers a fast and efficient way to gather up-to-date job information.

## Key Features

**Automated Job Search:** The script performs a job search query on Indeed through its API and returns job postings filtered by location (Chicago in the example).

**CSV Export:** Extracted data is saved into a structured CSV file, including fields like company name, job location, salary, and posting date.

**Error Handling:** The script includes basic error handling for failed requests and missing fields in the API response.

**Pagination:** The script is set to pull multiple pages of job listings, ensuring comprehensive data retrieval.

**Customizable Parameters:** The API query allows for adjusting job search parameters like location, search radius, and job type.

**Scalable for Larger Data Sets:** With some adjustments, the script can be scaled to gather large volumes of job data, making it suitable for data analytics and labor market research.

## Code Analysis

*The script is structured as follows:*

### Library Imports:

requests: Used to handle HTTP requests to the Indeed API.

json: For handling JSON responses.

csv: To store extracted data in CSV format.

### API Query Setup:

URL: The Indeed API endpoint for job searches.

Query Parameters: Define the job search filters such as job title (cybersecurity), location (Chicago), page number, job type, and more.

Headers: Authentication is handled via API key provided by RapidAPI.

### API Request & Response Handling:

A GET request is sent to the API, and the script checks for a successful response (HTTP status code 200).

The JSON response is parsed, and data is pretty-printed for debugging and understanding the structure.

### Data Extraction:

The relevant job data (company name, posting date, job ID, salary, etc.) is extracted from the response.

Missing fields are handled by validating that all required fields are present.

### CSV File Writing:

A CSV file is created, and the job data is written with appropriate headers.

Each job listing is written as a new row in the CSV file. Salary details are combined into a readable format.

Error Handling:

If the API response is unsuccessful or missing fields are detected, appropriate error messages are logged.

Example Code Enhancements

Pagination Loop: To gather more data across multiple pages of job listings, implement a loop that updates the `page_id` for each subsequent API call.

python

Copy code

```
# Loop over pages for paginated results
```

```
for page in range(1, 5): # Adjust the number of pages as needed
```

```
    querystring['page_id'] = str(page)
```

```
    response = requests.get(url, headers=headers, params=querystring)
```

```
    # Handle response as before
```

Date Conversion: Convert timestamps into a human-readable date format using Python's `datetime` module.

python

Copy code

```
from datetime import datetime
```

```
# Convert timestamp in milliseconds to a readable date format
```

```
job['pub_date_ts_milli'] = datetime.fromtimestamp(job['pub_date_ts_milli'] /  
1000).strftime('%Y-%m-%d')
```

## Resources/Research

RapidAPI Documentation: For understanding API usage, authentication, and query parameters for Indeed API.

RapidAPI Indeed Documentation

Python Requests Library: Essential for making HTTP requests and handling API responses.

Requests Documentation

JSON Handling: Tutorial on using Python's built-in json module.

Python JSON Guide

CSV Module in Python: For handling CSV file operations.

CSV Documentation

## Conclusion

This project automates job data extraction from Indeed using Python, focusing on cybersecurity roles. With customizable parameters, pagination, and error handling, the script provides a scalable solution for gathering and analysing job market data