

CSCI E-7 Graduate Project Checkpoint

Web browser: Google Chrome

Site to scrape: <https://www.jobbank.gc.ca/jobsearch/jobsearch>

Integrated development environment (IDE): Visual Studio Code (VSC)

Install the required Python libraries using the following command line in a VSC terminal (see Figure 1):

```
python -m pip install library_name
```

Site Inspection

Before scrapping the site, we should first inspect the site by checking how URL changes when different filters are used and how key objects on the page are written in HTML. Filter job search with at least two job categories to compare the URL change. See below.

Filter 01: Computer programmers and interactive media developers

Result 01:

<https://www.jobbank.gc.ca/jobsearch/jobsearch?searchstring=Computer+programmers+and+interactive+media+developers&locationstring=>

Filter 02: Web designers and developers

Result 02:

<https://www.jobbank.gc.ca/jobsearch/jobsearch?searchstring=Web+designers+and+developers&locationstring=>

Likewise, click onto different job cards to see the URL change. From this, we can see that each job posting is identified by an 8-digit unique identifying number (as highlighted). See below.

Job card 01:

<https://www.jobbank.gc.ca/jobsearch/jobposting/35151179?source=searchresults>

Job card 02:

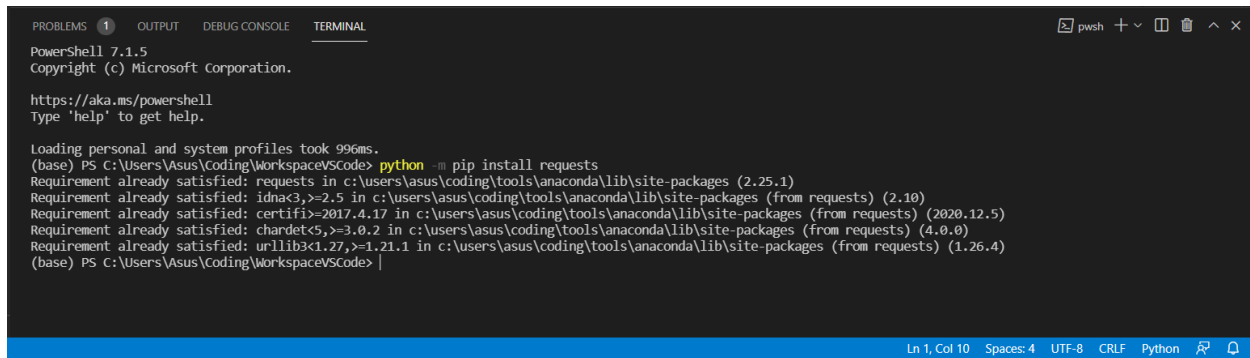
<https://www.jobbank.gc.ca/jobsearch/jobposting/35187140?source=searchresults>

Right-click on the page and choose [Inspect] or press [F12] will lead to a developer tool that show the HTML code of the page (see Figure 2). Use the selection tool on the top left corner

of the developer tool to locate the corresponding lines of HTML code of each job card. From this, we can see that all the job cards on the job search result page are grouped within the following `<div>` element and each job card is wrapped by an `<article>` element:

```
<div id="ajaxupdateform:result_block" class="results-jobs">  
    <article id="article-35240904">...</article>  
    <article id="article-35240905">...</article>  
    ...  
</div>
```

Figures



```
PowerShell 7.1.5  
Copyright (c) Microsoft Corporation.  
  
https://aka.ms/powershell  
Type 'help' to get help.  
  
Loading personal and system profiles took 996ms.  
(base) PS C:\Users\Asus\Coding\Workspace\VSCode> python -m pip install requests  
Requirement already satisfied: requests in c:\users\asus\coding\tools\anaconda\lib\site-packages (2.25.1)  
Requirement already satisfied: idna<3,>=2.5 in c:\users\asus\coding\tools\anaconda\lib\site-packages (from requests) (2.10)  
Requirement already satisfied: certifi>=2017.4.17 in c:\users\asus\coding\tools\anaconda\lib\site-packages (from requests) (2020.12.5)  
Requirement already satisfied: chardet<5,>=3.0.2 in c:\users\asus\coding\tools\anaconda\lib\site-packages (from requests) (4.0.0)  
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\asus\coding\tools\anaconda\lib\site-packages (from requests) (1.26.4)  
(base) PS C:\Users\Asus\Coding\Workspace\VSCode>
```

Figure 1 Package installation

Date: Nov 11, 2021

Student: Preesa Saewong (Harvard ID: 91556812)

The screenshot displays the Job Bank website interface on the left and a browser's developer console on the right. The website shows a search for 'Available jobs' with 85,329 results. Two job listings are visible: 'Social media community manager' at the Discovery Centre in Halifax (NS) and 'Reception clerk' at the GMF Centre Medical Du Parc in Granby (QC). The developer console on the right shows the HTML structure of the job listings, with the 'Social media community manager' listing selected. The console also displays the 'Elements' panel with the corresponding HTML tags and the 'Console' panel with the 'What's New' message.

Available jobs - Search - Job Bank

Government of Canada / Gouvernement du Canada

Job search | Career planning | Trend analysis | Hiring | Help

Job search tools

Available jobs - Search

Recent searches

85,329 results

New Remote work available

Social media community manager ✓ November 12, 2021

Discovery Centre
Halifax (NS)
Salary: \$20.76 hourly

New Share your resume

Reception clerk ✓ November 12, 2021

GMF Centre Medical Du Parc
Granby (QC)
Salary: \$15.00 to \$18.00 hourly (to be negotiated)

Filters Sort by Search

Elements

```
<div id="results-list-wrapper" class="results-list-wrapper col-md-12 noanim">
  <h1 property="name" id="wb-cont"></h1>
  <div id="results-list-content" class="results-list-content">
    <span class="skip-to-filters wb-init wb-disable-init visible-md visible-lg" style="display: none;"></span>
    <span class="skip-to-filters wb-init wb-disable-init visible-md visible-lg" style="display: none;"></span>
    <div class="results-nav"></div>
    <div class="clearfix"></div>
    <div id="ajaxupdateform:result_block" class="results-jobs">
      <article id="article-35240897">
        <a href="/jobsearch/jobposting/35240897?source=searchresults" id="ajaxupdateform:j_id_20_3_39:0:j_id_20_3_3b" class="resultJobItem"></a> == $0
        <span id="ajaxupdateform:j_id_20_3_39:0:favouritegroup" class="float job-action"></span>
      </article>
      <article id="article-35240896"></article>
      <article id="article-35240895"></article>
      <article id="article-35240894"></article>
      <article id="article-35240893"></article>
      <article id="article-35240892"></article>
      <article id="article-35240890"></article>
      <article id="article-35240889"></article>
    </div>
  </div>
</div>
```

Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessibility

Filter

Console What's New

Highlights from the Chrome 95 update

New CSS length unit authoring tools
Drag to change the unit value and select unit type from the dropdown in the Styles pane.

Hide issues in the Issues tab
Hide irrelevant issues so you can focus only on those issues that matter to you.

Figure 2 HTML code