**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**

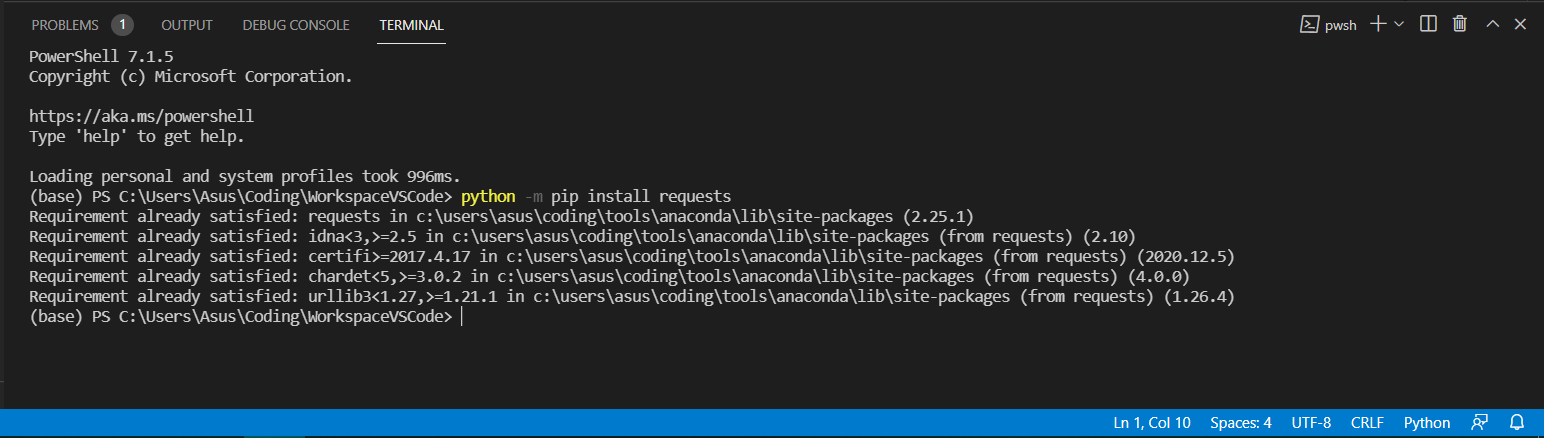


Figure 1 Package installation

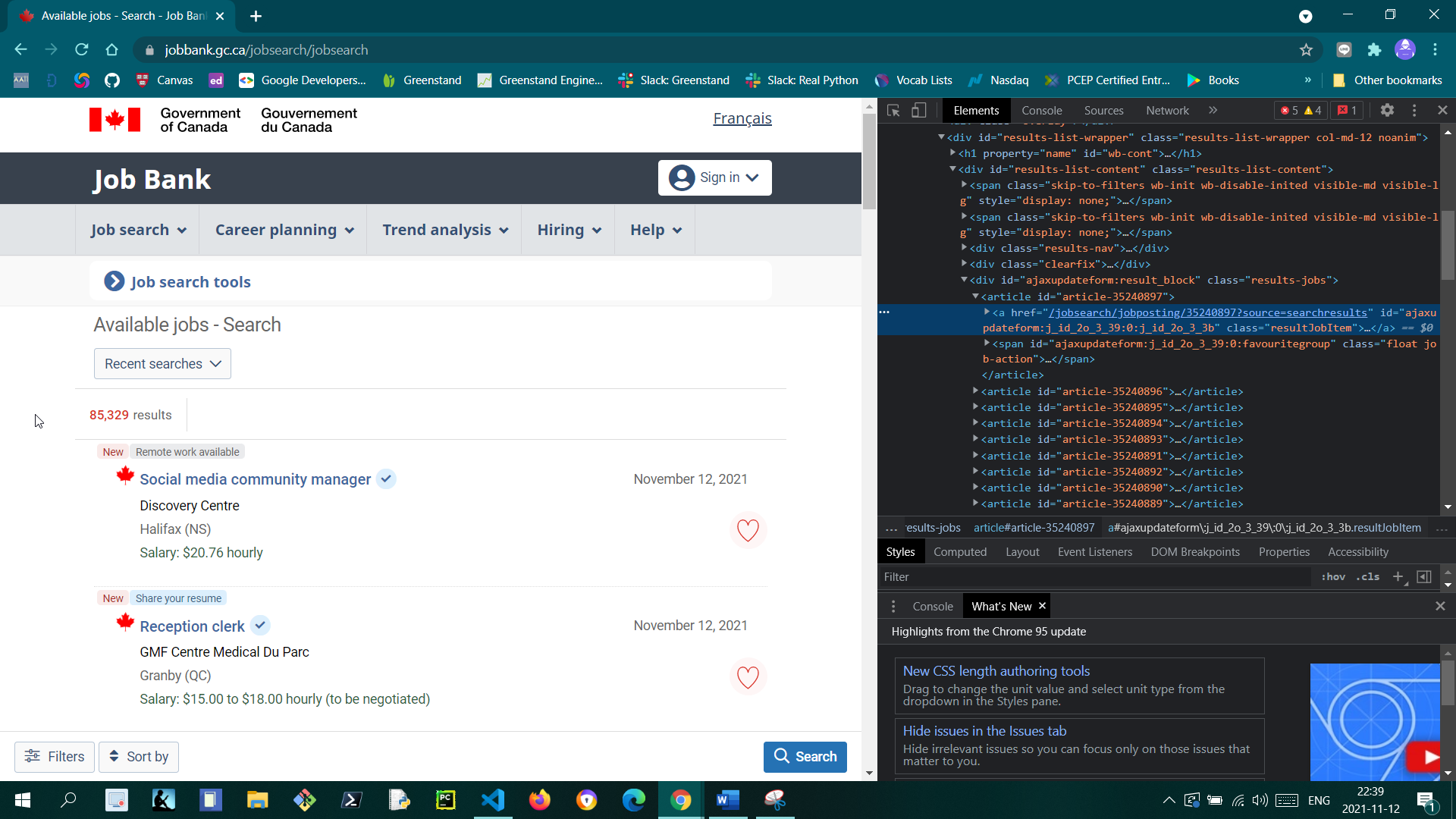


Figure 2 HTML code