# COMP 370 Homework 7 – Web Scraping Assigned Oct 25, 2023 Due Oct 30, 2023 @ 11:59 PM

In this homework, we’ll be using web scraping to collect the 5 trending stories from the Montreal Gazette. Our objective is to collect the title, publication date, author, and opening “blurb” (as seen in the screen capture here).

A screenshot of a computer

Description automatically generated

We want a script “collect\_trending.py” that does all the work (i.e., we don’t have to specify the trending articles one by one or in a list. The script goes, figures out which they are, and then grabs them off the website).

So to do this, you’ll have to write a scraper for two different page templates.

* To get the trending stories (and links to them), you’ll need to first scrape the homepage of Montreal Gazette (<https://montrealgazette.com/category/news/>)
* Then once you have links to the trending stories, you’ll need to scrape the key information off the article page itself.

collect\_trending.py is run as follows:

python collect\_trending.py -o trending.json

Such that trending.json has the format:

[

{

“title”: “article title”,

“publication\_date”: “date”,

“author”: “author”,

“blurb”: “blurb”

},

{

… article info

},

…

]

For both page templates, use cache-ing to avoid overly taxing the Montreal Gazette website.

## Submission Instructions

* Submit all your code in a zip file hw7.zip