

## Part 1:

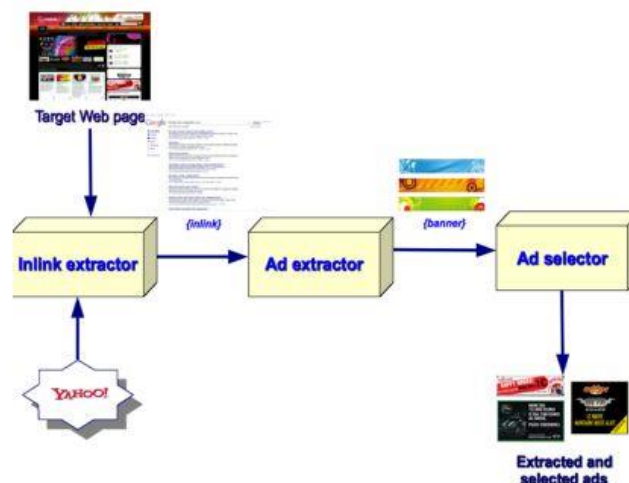
I did both the Fahrenheit to Celsius conversion, and vice versa, the sign (num), and also I did the gender exercise. My comment on the book is that it does not help with any suggestions if it fails to run as specified in the exercise. For instance, it will be very helpful to have the exercises in a separate page that is divided into two section or have a right or left bar, that tells you some suggestions to look for if the code did not work such as it may be a problem of indentations, make sure that you have the : symbol. It's very frustrating to spend time trying to figure out something you can do it online in less time if you're watching a video. My second comment is that the sign (num) exercise never worked with me, and tell now I don't know why. The gender exercise has a mistake I think that all the single " quotations should be double. It was frustrating at the beginning but when I figure it out, it worked fine. My file on github is called "chapter 7 exercises".

## Part 2:

For this week I selected this article:

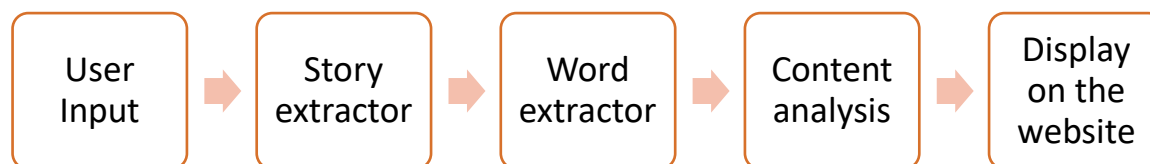
Vargiu, E., & Urru, M. (2013). Exploiting web scraping in a collaborative filtering-based approach to web advertising. *Artif. Intell. Research*, 2(1), 44-54.

The article introduce a collaborative system for web scraping that aims to suggest ads to a given Web page. The figure below demonstrate how the system works:



The system has three stages: (1) Inlink extractor, (2) Ad extractor, and (3) Ad selector. In the Inlink extractor module, the script is programmed to find peers pages on the internet and extract the first 10 Inlinks of a given page depending on Yahoo site explorer. The Ad extractor stage concerns with extracting banner ads from the peer pages relying on web scarping technique, they adopted scarping techniques to (i) access tags as object members; (ii) find out tags whose name, contents or attribute match some selection criteria; (iii) access tag attributes by using a dictionary-like syntax. They used Python BeautifulSoup for scraping the web. The output of this stage is a set of extracted banners and their urls which will go through the final stage the Ad selector. In the Ad selector, a random python file runs to select three suitable ad banners for the web page.

Building on this module, I organized my thoughts about my contact analysis project to go through these stages as shown in the figure below:



**User Input.** In this module, the python script is programmed to: (i) collect the user's keywords for the story, (ii) collect which newspaper the user wants to compare, (iii) collect the dates the user want to analyze (iiii) connect to NewsAPI and pass the all these information to the selected newspapers.

**Story extractor.** In this module, the python script will extract all the articles that are related to this story from the selected newspapers within the specified dates

**Word extractor.** In this stage the system will extract (i) headline title, (ii) the first 100 words (using the inverted journalism method) in the story

**Content analysis.** In this module, the script will calculate similar words (whether it is in lower case or in capital case). The script also will organize how many times the word appear in the title and how many times in the body.

**Display on the website.** The script will display the content analysis in a comparison table that shows the differences between stories for the user.