

logo.wmf

**Inside View**

**Version: 1.0 | Date: 11/09/2017**

Contents

[1 Inside View 3](#_Toc462408920)

[1.1 Inside View – Overview 3](#_Toc462408921)

[1.2 Prerequisites 4](#_Toc462408922)

[1.3 Installations & Operations 5](#_Toc462408923)

[1.3.1 Connection Configure Changes 5](#_Toc462408924)

[1.3.2 Informatica Configuration Changes. 6](#_Toc462408925)

[1.3.3 Create template Changes. 7](#_Toc462408926)

[1.4 Outputs 8](#_Toc462408927)

[1.5 Limitations 9](#_Toc462408928)

# Inside View

## Inside View – Overview

Inside View is a text analytics crawler that can be used to extract content from web pages and perform sentiment analysis and topic modelling.

This tool significantly reduces the amount of expertise required to extract the contents from the web pages and allows to do sentiment analysis with minimal effort.

## Prerequisites

Machine requirement:

1. Windows 7 and above
2. 64 bit processor preferably (Solution will also work on 32 bit)
3. Min RAM size – 4 GB

Infrastructure:

1. Internet connectivity to download the software’s
2. Admin privileges to install the softwares.

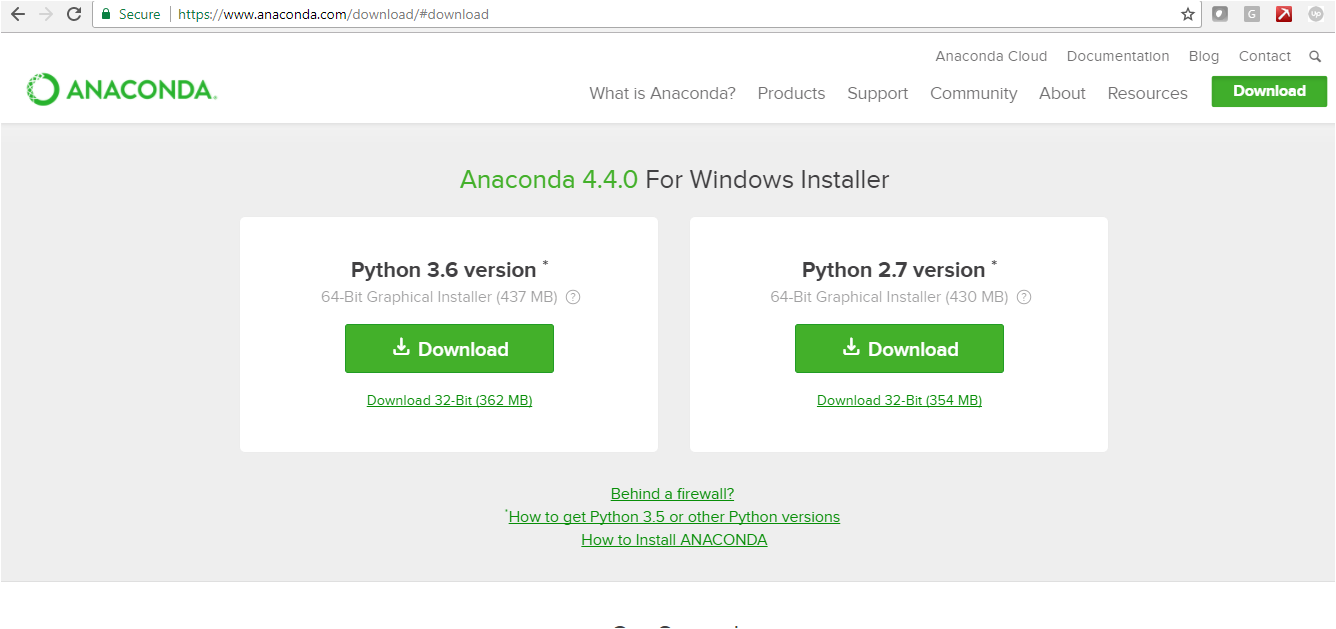
Software Requirements

1. Anaconda python latest version
   1. Python Version 2.7x (preferably)
2. Google Chrome
3. Chrome driver for selenium (for corresponding processor).

## INSTALLATIONS AND OPERATIONS

**Anaconda and Python Installations**

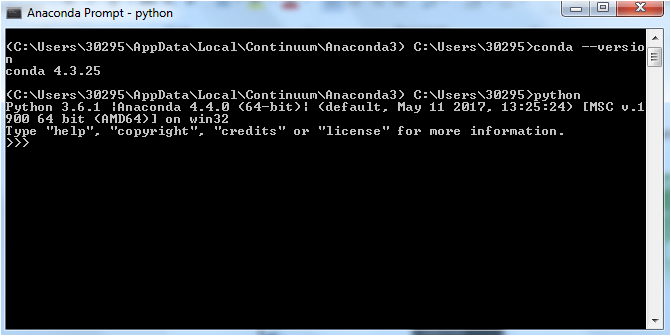
1. Anaconda python Installation:
   1. Go to this link for downloading python via anaconda distribution- <https://www.anaconda.com/download/#download>



* 1. Select the desired python version to download (Python 2.7 preferred)
  2. Run the .exe file to install python
  3. Once the installation is done, check for “anaconda prompt”



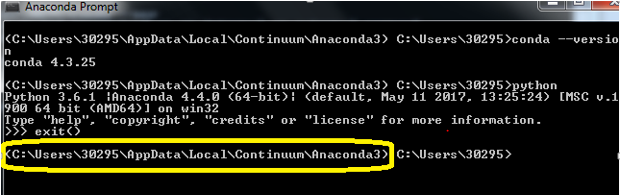
* 1. Open the anaconda prompt and enter the command “python”. Press enter to check the Python and Anaconda versions. If the python is installed correctly, then proceed with step “l”



* 1. If the response throws an error, you need to manually set the path of python to the system path.

**Setting up Environment Variable (This is required only if step “e” fails)**

* 1. To set the python system path, navigate to the “Anaconda Python folder”. Anaconda python path can be seen in the “anaconda prompt” as shown below

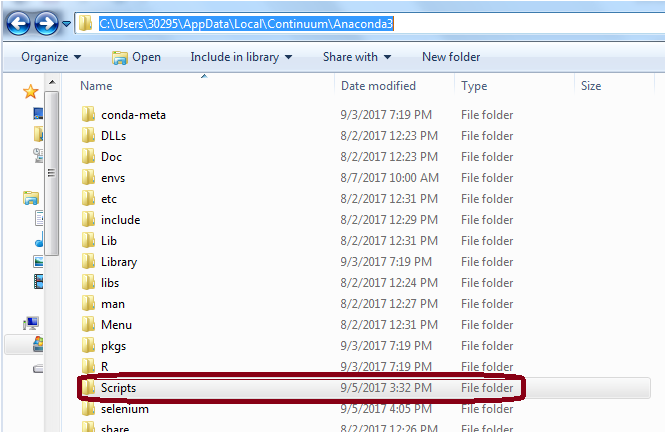


* 1. Make a note of the folder path ( the path which is marked in the above image ). Look for the folder named “scripts” and make a note of this folder path as well.

For example:

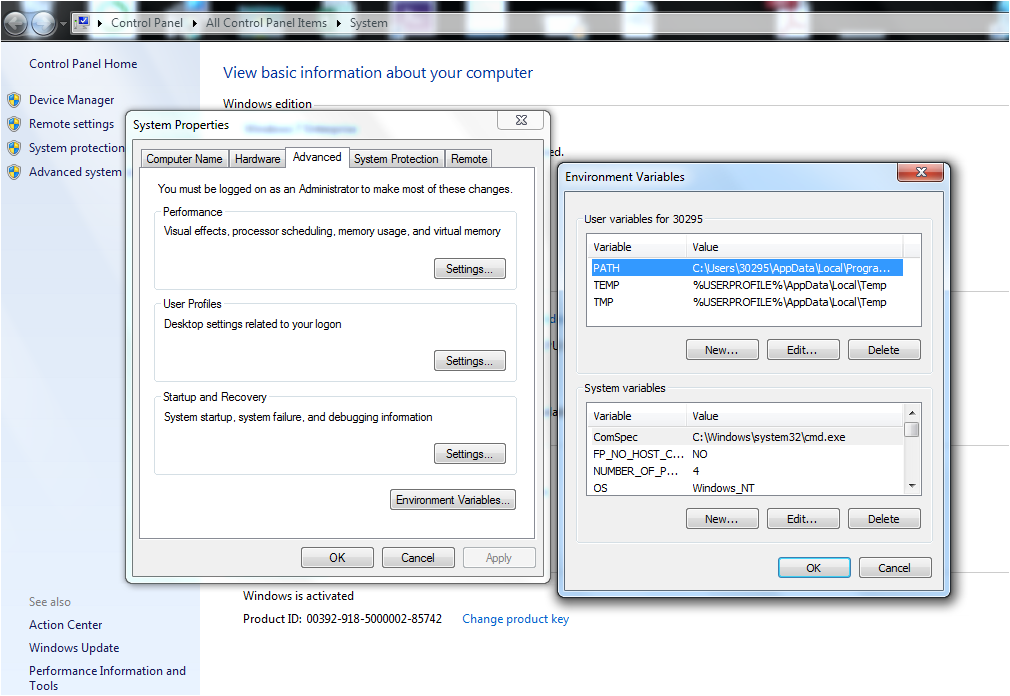
**C:\Users\30295\AppData\Local\Continuum\Anaconda3\Scripts –** path for script

**C:\Users\30295\AppData\Local\Continuum\Anaconda3 –** path for python

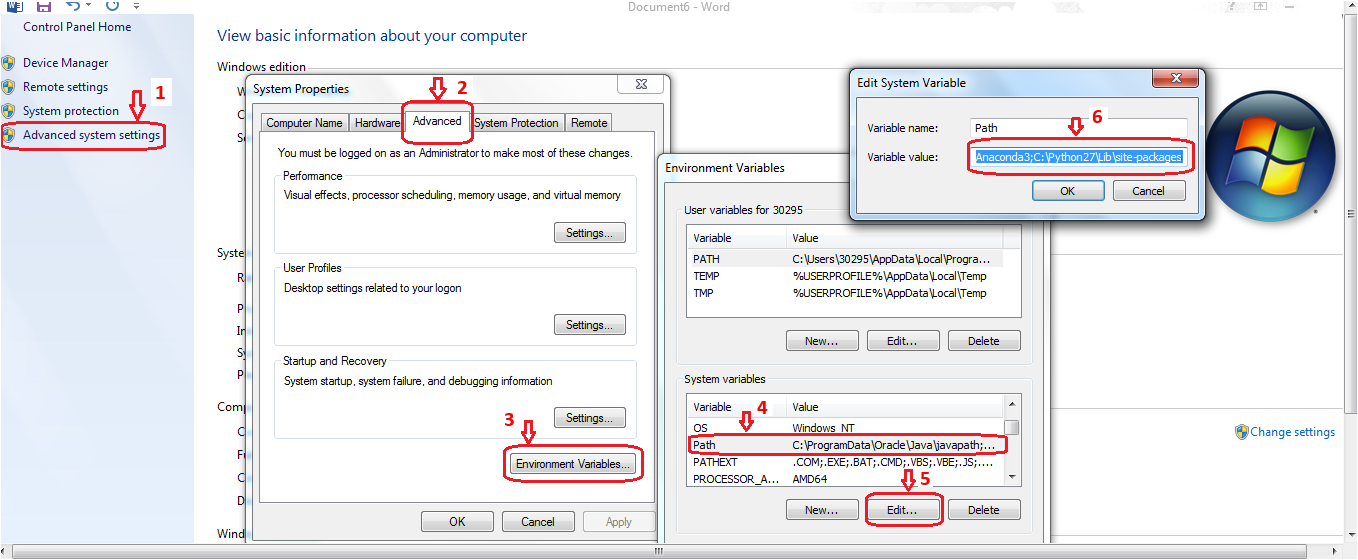


* 1. Navigate to set the system variables as follows:

Control Panel >> All Control Panel Items >> System >> Advanced system Settings >> Advanced (tab) >> Environment Variables



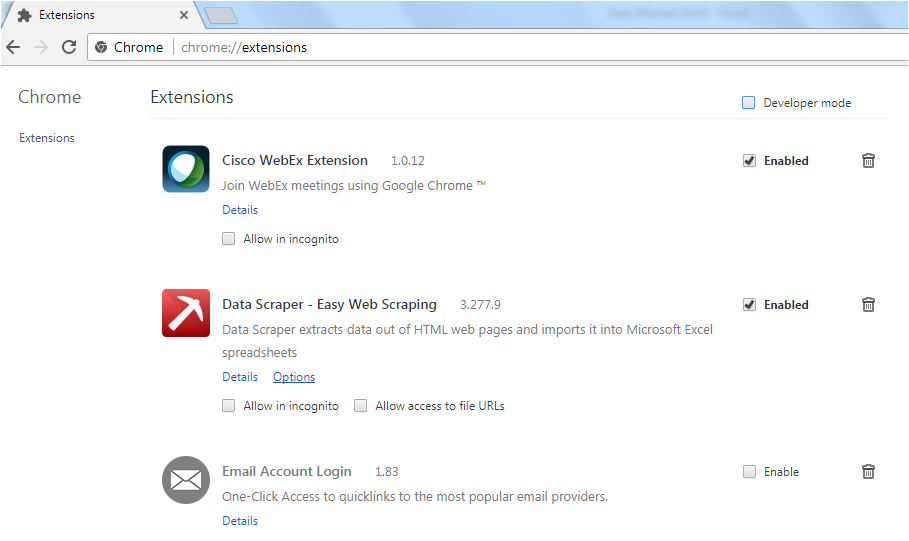
* 1. Copy the “path for script” and “path for python” as mentioned above and paste it in the system variables.
     1. Navigate to “Path” in the system variables window, select “path” and click on edit button
     2. Now append both the paths (“path for script” and “path for python”) to the end, separated by a semi-colon (;) and click ok for all the windows.



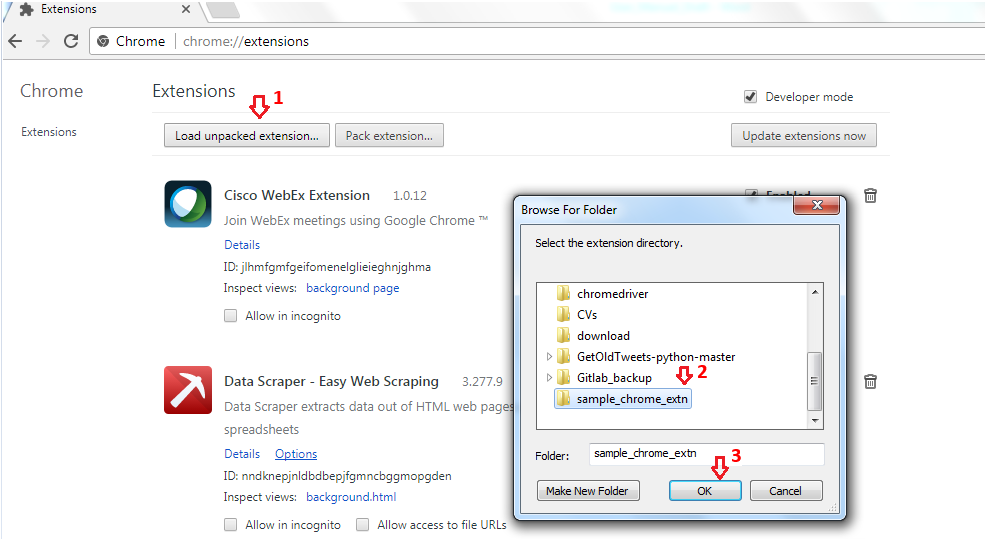
* 1. Repeat step “e” to check for the python and anaconda versions.

**Installing the chrome plugin for on-click event**

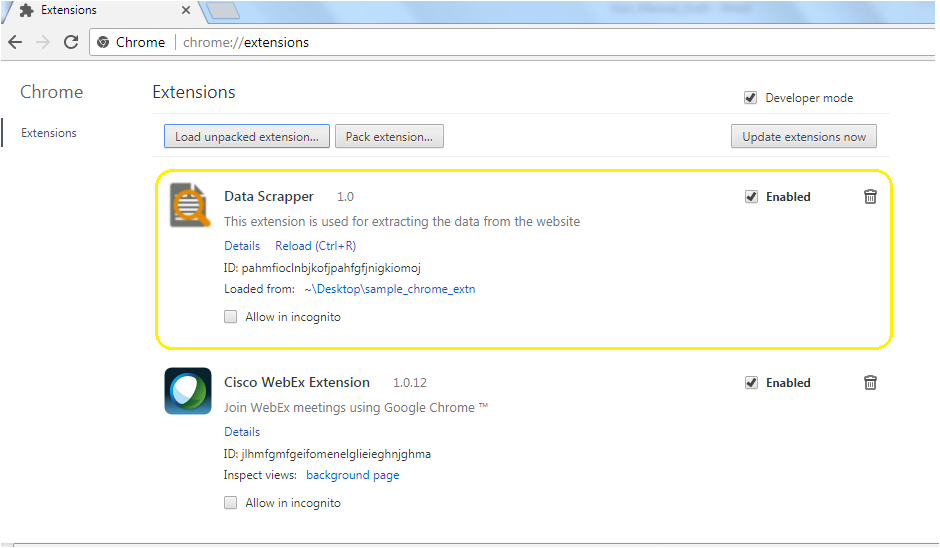
1. As part of this software package, we have provided a chrome plugin to handle on-click events on the chrome browser.
2. To add this plug-in to chrome, open chrome extensions window by typing the following url in the chrome browser : **chrome://extensions/**



1. Enable the “Developer mode” check box.
2. Click “Load Unpacked Extension” to add the chrome plugin binding which we have provided. Navigate to the folder where the binding is saved as shown below.
3. Select the entire folder and click OK to add the plugin.



1. After adding the plugin, it will be available in the chrome extensions window as shown below.

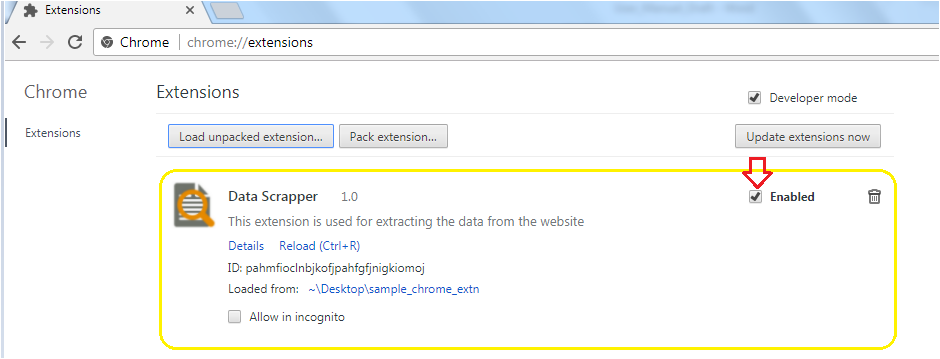


**Enabling the plugin to select the desired element**

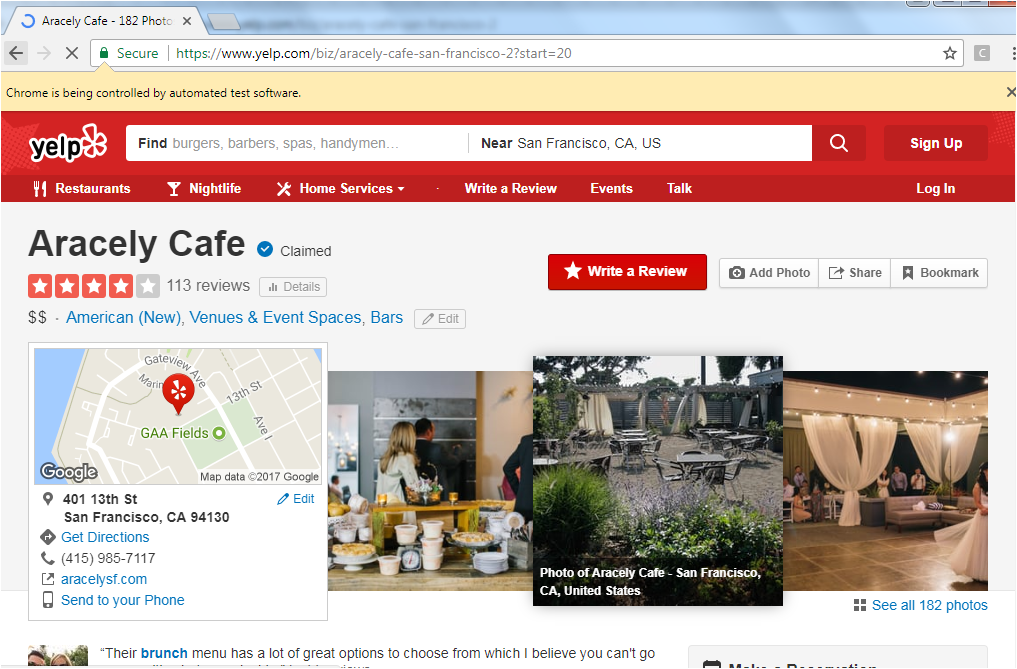
1. Go to chrome extensions by typing the following url in the chrome browser :

chrome://extensions/

1. Enable “Data Scrapper” as shown below

* 

1. Open the web page for which the data needs to be extracted in the browser

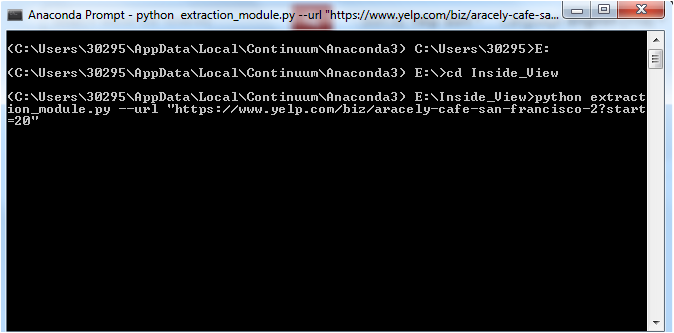


1. Navigate to the folder where the data extraction code is available which we have provided.
2. Open Anaconda terminal from the folder where the program in available.
3. Execute the following script by passing the required url for which the date in to be extracted as argument.(copy the url from the browser and pass it as argument)

Command: python extraction\_module.py --url **“url\_input”**

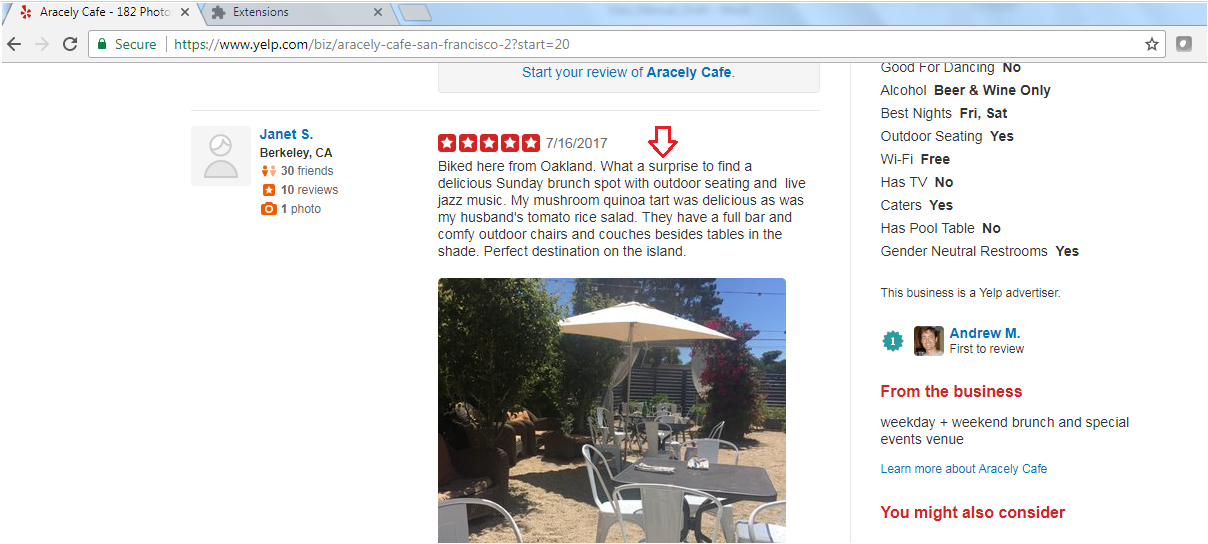
For example:

**python extraction\_module.py --url “https://www.yelp.com/biz/aracely-cafe-san-francisco-2”**

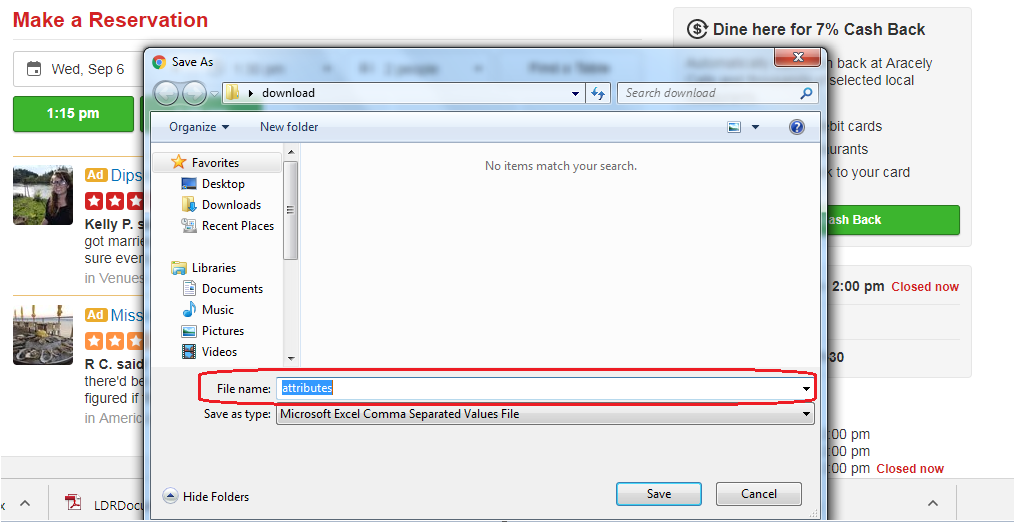


1. Executing the above script will invoke the extraction module engine.
2. Click on the desired element for which the data needs to be extracted

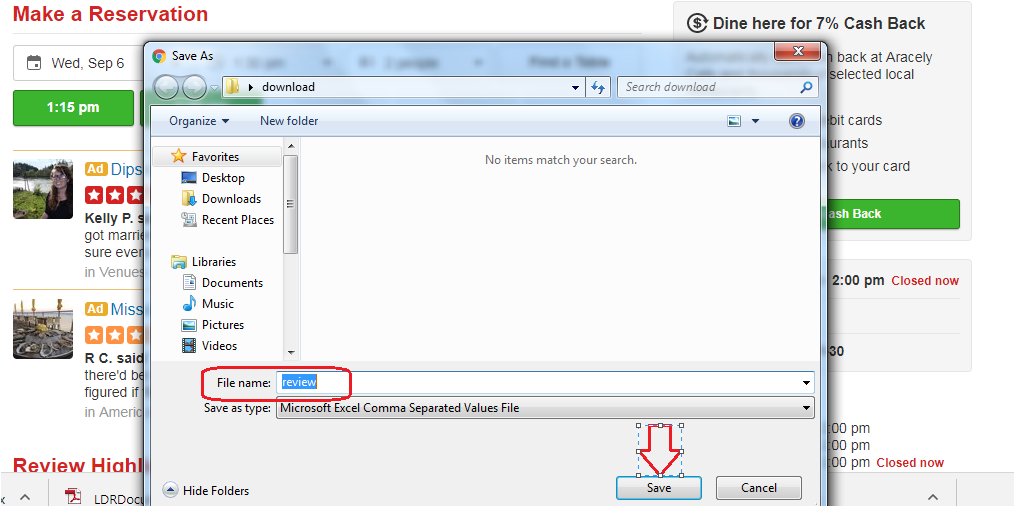
For example: Here we are clicking on the “product review” as shown by the red arrow.



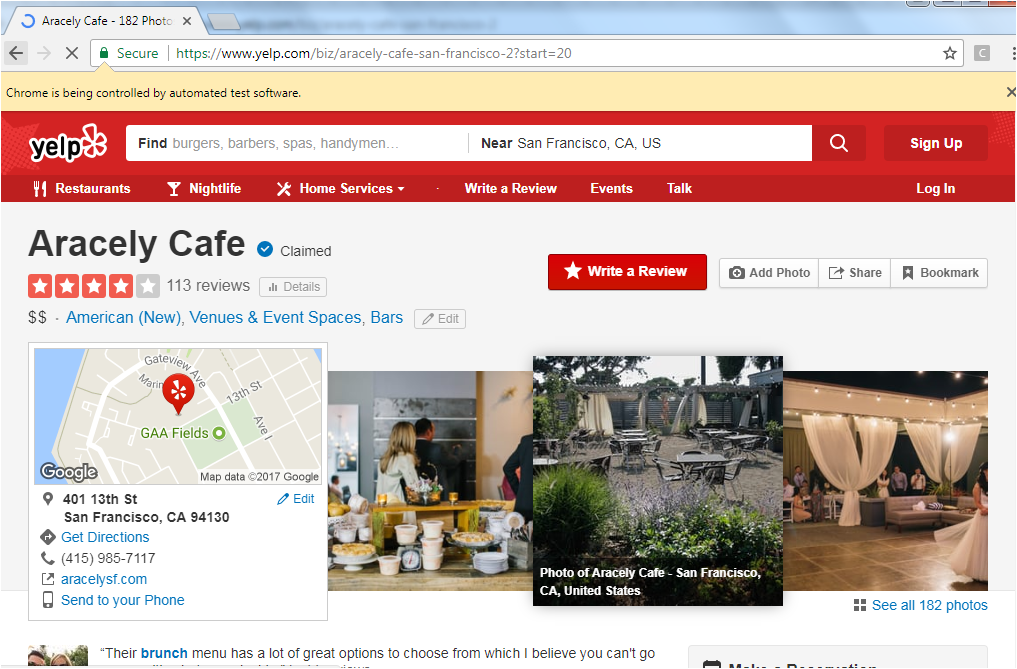
1. The attributes of the element will be stored in as CSV file.



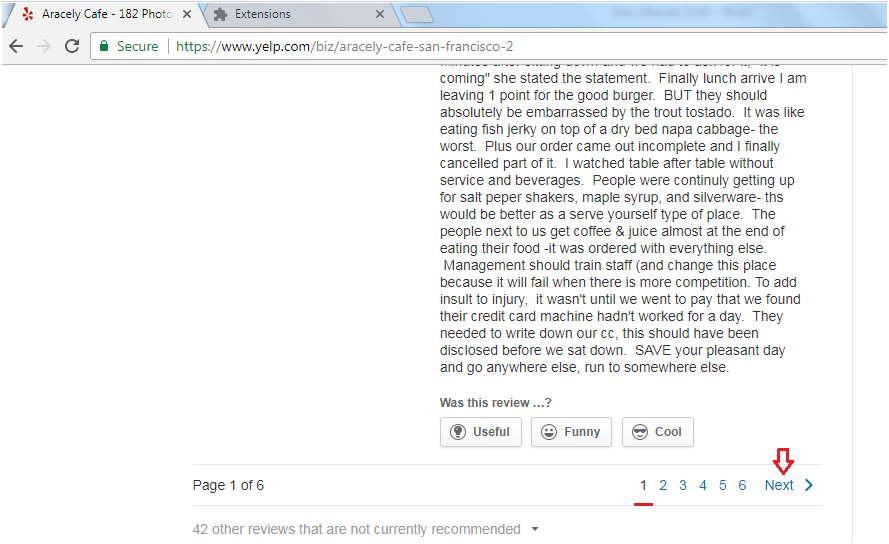
1. Rename the file as “review” and save.

* 

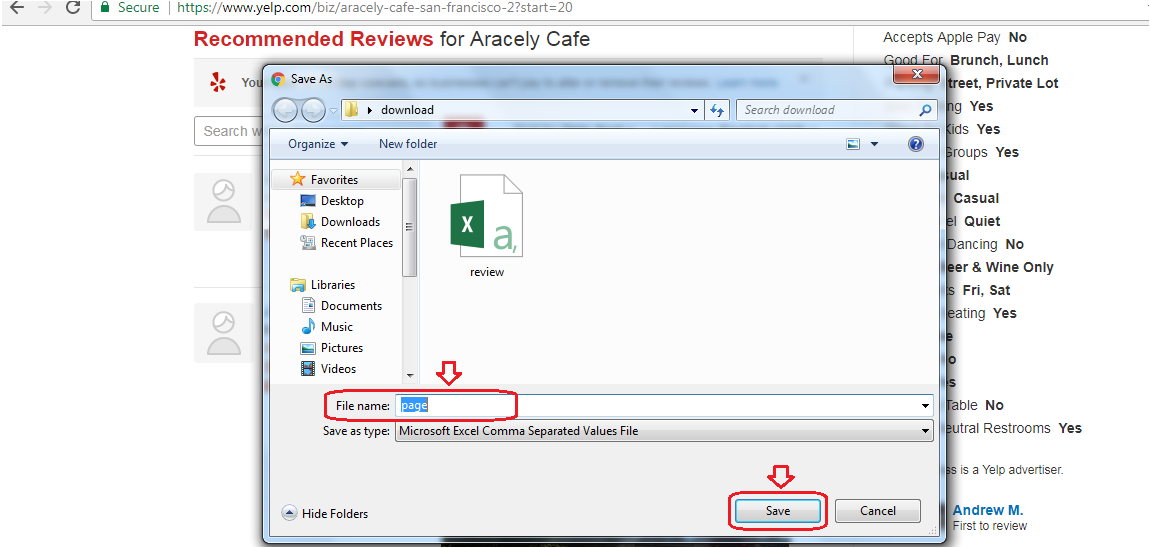
1. A separate window for data extraction will open (as shown below). \*\* D**o not disturb this window till the extraction gets complete.**



1. For extracting the content from the subsequent pages (if needed), scroll down to bottom of the page and click next page element as shown below.



1. **\*\*Note: Select the “Next” element as shown above for extracting content from subsequent pages. This element will not be same for all other website. Don’t select the individual page elements.**
2. Select the “Next” element and save the following file by renaming it as “page”.



1. Data will be extracted from subsequent pages till the final page. Final output is saved as “extracted\_data\_review.csv”.

**Performing sentiment and topic modelling**

1. Navigate the Anaconda prompt folder containing the python engine for performing sentiment and topic modelling.
2. Execute the below mentioned command for performing the sentiment analysis and topic modeling

Command: python Sent\_topic.py --file **"*path to the extracted\_data\_review.csv* "**>> **“file\_name”**.csv

Example:

python Sent\_topic.py --file "C:\Users\30295\Desktop\extracted\_data\_review.csv" >> output.csv

1. The output for sentiment and topics are saved in output.csv file.

