Assignment 6 Data Extraction from Social Media

Method · November 2019

CITATIONS
0

READS
29

2 authors:

Nimat Ullah
Vrije Universiteit Amsterdam
6 PUBLICATIONS **1** CITATION
SEE PROFILE

Jan Treur
Vrije Universiteit Amsterdam
828 PUBLICATIONS **8,231** CITATIONS
SEE PROFILE

Assignment 6 Data Extraction

Assignment Objectives

- To practice dealing with available data extraction and visualization tools.
  - Learn how to extract data of your choice from web.
  - How to visualize and analyze available data with some visualization tool.

In this assignment you are free to personalize/use any available software/tool for data extraction or visualization. Below are links to some of the available (mostly trail versions) tools.

Expected Outcome

As an outcome of this assignment, you are asked to:
1. Extract data of your choice from the web,
2. Visualize the data, and
3. Explain your analysis of the data in a short report (in one page).

Point of departure

You need to decide which data extraction and visualization tool are you going to use. Make sure that the data visualization tool supports the format of data that the extraction tool offers you.

Note: The tools listed below are just for laymen and not obligatory to use for this assignment. You can choose or code on your own choice depending on your data requirements.

1. https://socioviz.net/SNA/eu/sna/login.jsp. (for twitter hashtags)
2. https://netlytic.org. (for Twitter, Instagram, YouTube, Cloud Storage, Text Files and RSS).
3. https://archive.codeplex.com/?p=nodexl (for windows users, supporting various social media sites)
   a. https://www.smrfoundation.org/nodexl/installation/ (installation procedure for nodeXL)
4. https://www.parsehub.com (for extraction of data from web pages)
5. https://www.scrapestorm.com/ (100 rows of data per day can be imported in trail version from any website)
6. http://webextract.net/download_best_web_scraper_for_free.aspx (Can only be used by windows users for data extraction from any URL)
7. http://www.netminer.com/main/main-read.do (can only be used by windows users for Instagram as well as Facebook, YouTube, Twitter and various other sources)
8. www.fminer.com
9. http://www.tweetstats.com/graphs/imrankhanpti (twitter online analytics)
10. https://unionmetrics.com/free-tools/twitter-snapshot-report/ (Twitter hashtags and Instagram checkup)
11. www.newprosoft.com (can be used for extraction of data and images from any website automatically)
12. https://www.webscraper.io (also available as Google Chrome extension and can be used for data extraction from any website)
13. http://socialmention.com/# (searches the mention of a given phrase/hashtags in blogs, microblogs, bookmarks, images, videos etc)
14. https://commoncrawl.org/the-data/ (it has data that's stored in WARC file format)

- Most of the above tools are listed in:
  - https://www.octoparse.com/blog/top-30-free-web-scraping-software

    OR
  - https://www.kdnuggets.com/2015/06/top-30-social-network-analysis-visualization-tools.html

- If you're planning to visual & analyze your extracted data in Gephi then go through the following link to know about the formats supported by Gephi.
  - https://gephi.org/users/supported-graph-formats/