International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

A Novel Application of Data Analytics in Data Journalism

Ram Mohan Rao P, Hima Keerthana,
*(Associate Professor and Head, CSE Department, Sphoorthy Engineering College,Hyderabad
Email: rammohan04@gmail.com)
** (CSE Department, Sphoorthy Engineering College, Nadergul
Email: abcdef@yahoo.co.uk)

1. ABSTRACT-

Today we live in a digital age, where numbers can be used to express virtually everything (and almost anything). Data journalism is a type of journalism in which reporters create stories using massive datasets. This reflects the growing importance of numerical data in the production and dissemination of information, as well as the growing connection between journalists and professions like design, computer science, and statistics. Through appealing info graphics, data journalism may assist a journalist in telling a complicated narrative. Data may be the source of data journalism, or it can be the instrument used to tell the story—or both. It should be handled with suspicion, just like any other source, and we should be aware of how it might influence and limit the tales that are generated using it, just like any other instrument. In this project we apply web scraping techniques to implement data journalism and help reporters and media houses to harvest the power of scraping in extracting the facts from the data, attract, engage and promote innovative solutions for their audience.

Keywords- Data Journalism, Web Scraping, Info graphics.

2.INTRODUCTION-

Journalism is one of the important activities of human life which deals with publishing facts, ideas, and events occurring in society and keeps the public informed about the issues around them. However, in today's digital world the digital presence of the public is inevitable and eventually created a digital society. Lot of things do happen in the digital world and there is a need to bring the facts, ideas and events to the public and keep them informed which is a very big challenge with the scale of data being generated today.

Data journalism is an innovative concept of online journalism where web scraping is applied to extract data from different places, filter the same, segregate them and publish in the form of legitimate news. Data journalism plays a vital role in today's digital world because the presence of humans is more digital than physical. Example: data scraping can be applied on e-commerce platforms and generate a report on buying standards of public in a country.

We will be scraping data from different web sites based on certain theme or interesting concepts, using a Data Web Scraping which can give a visualized or excel sheets overview of topic over a desired period, from collection of sites in which the report is mentioned by their websites. We will be scraping data from different web sites based on certain themes.

Using Web Scraping which can give a visualized or excel sheets overview of topic over a desired period, from collection of sites in which the report is mentioned by their websites. The person who is in research with the topic can get knowledge and he can also improve their research by our scraped data and can be able to take decisions smoothly. The extracted data from the web is either for personal use by the scraping operator, or to reuse the data on other websites

3. LITERATURE SURVEY-

The aim of a study has two main objectives: Firstly, the study aims to discuss some hurdles and challenges faced by the data journalism community in Latin America which limit the extent to which data journalism can produce reporting that reaches a mainstream audience, as opposed to being limited to specific niches. There are some challenges, they may include issues such as lack of resources, limited access to data, or legal restrictions on freedom of the press. Secondly, the study aims to propose innovative and resilient approaches to overcome these obstacles. There are ways to surmount the challenges faced by data journalism in Latin America and increase the reach and impact of data reporting. The approaches include strategies such as developing partnerships with other media outlets, using alternative data sources, or using data visualization tools to make reporting more engaging and accessible.

4. IMPLEMENTATION-

a. Web scraping is a powerful tool for journalists looking to gather data from websites and online sources. Here is some example code for web scraping using the Python programming language:

```
import requests
from bs4 import BeautifulSoup
# set the URL of the webpage to scrape
url = "https://example.com"
# send a request to the webpage and get the HTML
content
response = requests.get(url)
html_content = response.content
# parse the HTML content using BeautifulSoup
```

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

```
soup = BeautifulSoup(html_content, 'html.parser')
# extract the data you need from the webpage
# for example, find all the links on the page
links = []

for link in soup.find_all('a'):
links.append(link.get('href'))
# write the data to a file or database
# for example, write the links to a CSV file
import csv
with open('links.csv', 'w', newline='') as csvfile:
writer = csv.writer(csvfile)
for link in links:
writer.writerow([link])
```

This code sends a request to a webpage and retrieves the HTML content. It then uses BeautifulSoup to parse HTML and extract the data needed. In this example, it finds all the links on the page and writes them to a CSV file.When using web scraping for data journalism, it is important to ensure that you are not violating any website's terms of service or copyright laws. Always check the website's policies before scraping their content and consider using an API or contacting the website's owners for permission.

b. The layout code of web scraping for data journalism can vary depending on the website and the specific data you want to extract. However, here is a general outline of the steps involved:

- Identify the website and the page(s) you want to scrape.
- Inspect the page(s) to determine the HTML structure and tags that contain the data you want to extract.
- Use a web scraping library such as Beautiful Soup or Scrapy to parse the HTML and extract the desired data.
- Clean and transform the extracted data as necessary, using techniques such as regular expressions, string manipulation, or pandas data-frames.
- Store the extracted and processed data in a structured format such as CSV, JSON, or a database.

```
import requests
from bs4 import BeautifulSoup
# send a request to the webpage and get the HTML content
url = "https://www.example.com/news"
response = requests.get(url)
html_content = response.content
# parse the HTML content using BeautifulSoup
soup = BeautifulSoup(html_content, 'html.parser')
# extract the data you need from the webpage
headlines = []
for headline in soup.find_all('h2', {'class': 'headline'}):
headlines.append(headline.text.strip())
# write the data to a CSV file
import csv
with open('headlines.csv', 'w', newline='') as csvfile:
writer = csv.writer(csvfile)
for headline in headlines:
writer.writerow([headline])
```

This code sends a request to a news website, extracts the headlines from the HTML using BeautifulSoup, and writes them to a CSV file. The specific HTML tags and attributes used may differ depending on the website and the data you want to extract.

5. Results-

5.1 Drop down list

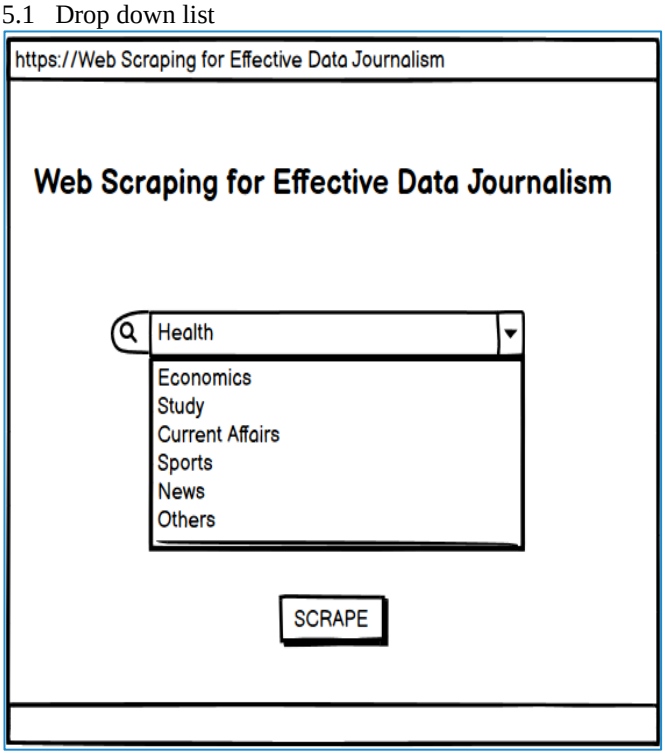

Fig. 1 The above diagram shows the UI to select the theme of data journalism from list of various drop-down news topics.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

5.2 URLs based on the given information

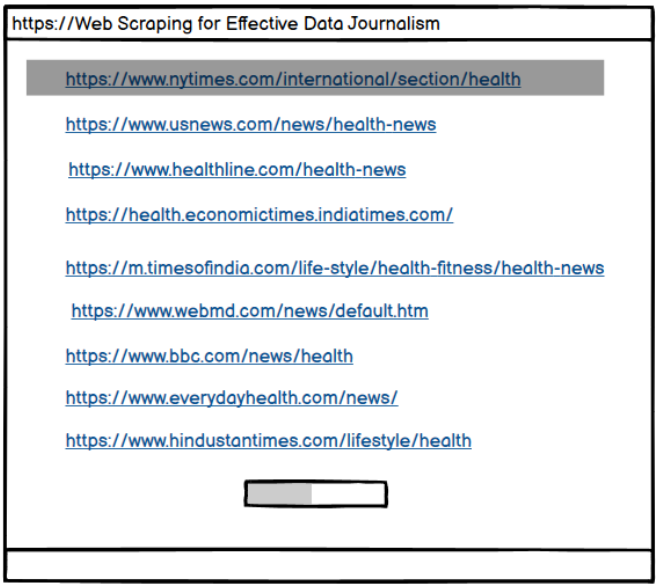

Fig.2 shows the result of the selected theme. The given information is fetched and provided by no.of URL's

5.3 Download the content

Fig.3 depicts us about the select URL content which is to be downloaded.

5.4 Extraction of the content

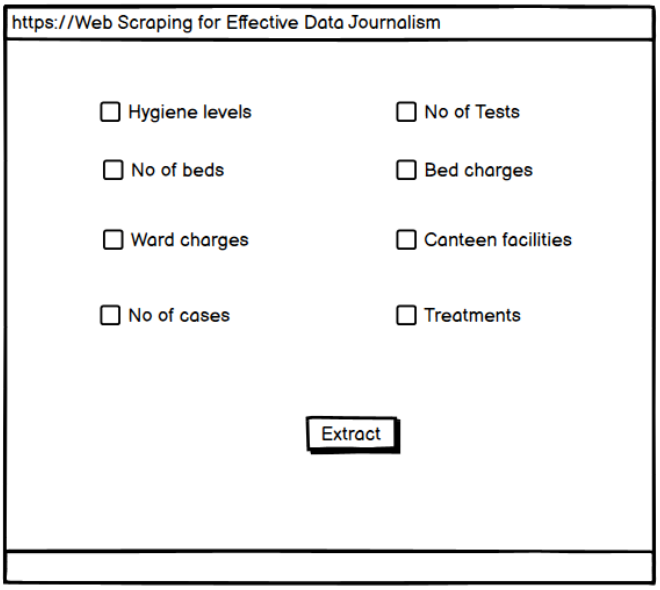

Fig.4 shows the result of the fields from the website provided to the user to choose based on their interest.

5.5 Stores in .csv files

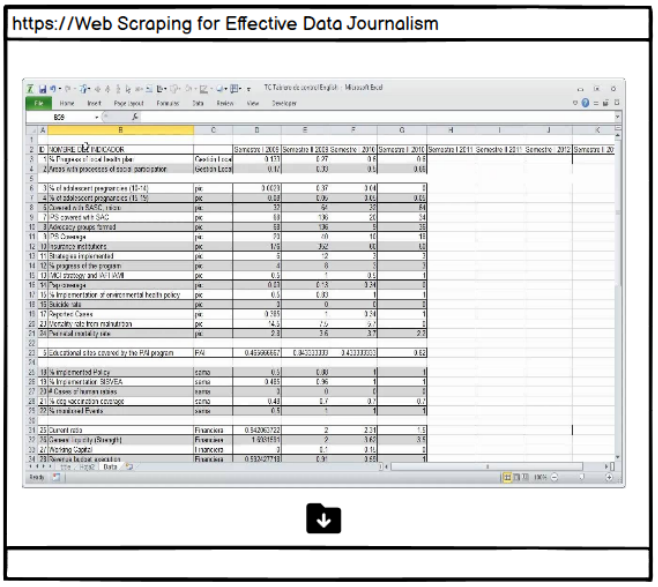

Fig.5 show that the information is stored in the form of .csv files to extract insights and create a new article using data analytics.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

6. CONCLUSION-

Data journalism is a modern method of aggregating useful information from the web, curating the information and extracting useful insights which can be treated as a news item. The same can be shared by all stake holders. Rapid growth of digital applications has led to creation of large volumes of data in the public domain which can be used to create useful news.

7. ACKNOWLEDGEMENT-

This work has been carried out as part of our academic project to be submitted to the university. In this project work, we got the guidance and all inputs from our internal guide Mr. Ram Mohan Rao P and we are thankful to our guide for his constant support and encouragement without which, the paper could not be completed.

8. REFERENCES-

[1] Alfter, Brigitte, and Stefan Cândea. 2019. "Cross-Border Collaborative Journalism: New Practice, New Questions." Journal of Applied Journalism & Media Studies 8 (2): 141–149
[2] Belair-Gagnon, Valerie, and Allison J. Steinke. 2020. "Capturing Digital News Innovation Research in Organizations, 1990–2018." Journalism Studies 21 (12): 1724–1743.
[3] Boczkowski, P. J. 2004. Digitizing the News: Innovation in Online Newspapers. Inside Technology. Cambridge and London: MIT Press.
[4] Carson, Andrea, and Kate Farhall. 2018. "Understanding Collaborative Investigative Journalism in a 'Post-Truth' Age." Journalism Studies 19 (13): 1899–1911.
[5] Hendrickx, Jonathan, and Ike Picone. 2020. "Innovation Beyond the Buzzwords: The Rocky Road Towardsa Digital First-Based Newsroom." Journalism Studies 21: 2025–2041.
[6] Pavlik, John V. 2013. "Innovation and the Future of Journalism." Digital Journalism 1 (2): 181–193. Posetti, Julie. 2018. "Time to Step Away from the 'Bright, Shiny Things'? Towards a Sustainable Model of Journalism Innovation in an Era of Perpetual Change."
[7] Lewis, Seth C., and Nikki Usher. 2014. "Code, Collaboration, and The Future of Journalism: A Case Study of the Hacks/Hackers Global Network." Digital Journalism 2 (3): 383–393.
[8] Mutsvairo, Bruce. 2019. "Challenges Facing Development of Data Journalism in Non-Western Societies." Digital Journalism 7 (9): 1289–1294.
[9] Paulussen, Steve. 2016. "Innovation in the Newsroom." In The SAGE Handbook of Digital Journalism, edited by Tamara Witschge, C. W. Anderson, David Domingo, and Alfred Hermida, 1st ed., 192– 206. London: SAGE Publications Ltd.
[10] Verganti, Roberto. 2009. Design Driven Innovation: Changing the Rules of Competition by Radically Innovating What Things Mean. 1st ed. Milan: Harvard Business Review Press.
[11] Appelgren, Ester, Carl Gustav Lindén, and Arjen van Dalen. 2019. "Data Journalism Research: Studying a Maturing Field Across Journalistic Cultures, Media Markets and Political Environments." Digital Journalism 7 (9): 1191–1199.
[12] https://www.icij.org/inside-icij/2018/09/web-scraping-how-to-harvest-data-for-untold-stories/
[13] Data Journalism Beyond Technological Determinism (tandfonline.com)
[14] https://balsamiq.com/givingback/free/