Universal Content Curator: An Improvised Web Scraping for the Online Content

Bhavesh K. Sutar[1] and Sudeshna Roy[2]

Student, Department of Information Technology[1]
Professor, Department of Information Technology[2]
Bharati Vidyapeeth' Institute of Management and Information Technology, Navi Mumbai, India

Abstract: *We are going in the direction of the Web global and IOT with the rate of light. Personalization and Recommendation systems are the pillars of internet. Now a day's users need to look the applicable content material on internet in minimal quantity of time. It's hard to make an efficient system which searches the relevant results on a click. To display the curated results and recommends the high-quality search results as consistent with user's interest is an ordinary task, and receives more complex while we're going to apply on web information. In this paper, a method and proposed model is mentioned that's technically recommends the relevant records as constant with his/her interest. The technique is used to retrieve records from web and refine the outcomes. Some present personalized looking strategies and models also are compared as literature survey.*

Keywords: Content curation, Web Scraping, Historical Data, Online Content

I. INTRODUCTION

Content curation is the approach of finding and collecting online content and providing the high-quality portions. Unlike content marketing, curation does no longer contain developing your very own content material.

Web searching is the need of contemporary technology and these days there are heaps of users. Most of the net users have described the net surfing as a hobby and the location of learning. The scope of net is unfold round the sector in every location like marketing, research, shopping. For the smooth net usage, there are rigorous efforts of internet crawlers and the searching strategies.These crawlers and searching techniques had been helping and supporting people to have a take observe the web. The task isn't always so easy. It's a technique wherein some strategies are implemented to present the personalized results. This is a critical and challenging task to achieve, which is 100% correct and matches the user's expected requirement.

It's vital for content entrepreneurs and users to get old, up to date and relevant data in much less time. The content curation comes into the picture where users and marketers manually collect already present data from web and create a new content based totally on present data, however gathering data from different web sources and create new content is really hard work for users and content marketers. The data mining strategies have been used to search and mine data, however now internet data isn't always restrict to structure data. User wants to search the applicable data and records in a minimal amount of time. Curation is the term used to fetch the data that is relevant to the user and as per user's interest So, "what if we are able to do these tasks using technology?"

The new method which called "web scraping" programmers able to retrieve any type of data within a second based on programming languages and automation. Web scraping is an automated technique to achieve huge amounts of data from websites. Most of this data is unstructured data in an HTML layout that is then converted into structured data in a spreadsheet or a database in order that it could be used in numerous applications. Here **Fig 1**, the model discussed the "content curation" process using "web scraping" and programming.

II. PROBLEM DEFINITION

It's been determined that Web Scraping is the most popular approach for accessing any type of data from the web. Generally, users and digital marketers doing The Content Curation manually, they're constantly spending their precious time on searching historic and updated data. But due to a massive data available on the web, it's so tough to find the relevant information. Though, relevant data to user is a tough challenge because of massive amount of missing data.

Though, it isn't always successful to recognize the user's interest. Therefore, developing an application that allows you content curation method automatically "Web Content Curation".

I.e. performing web scraping approach with programming language to make software that may gather the information as per user's interest from popular and trusted sources, additionally do filtration process automatically on accumulated data and provide relevant information within less amount of time.

III. OBJECTIVE / SCOPE

My long term objective with this research is determining how much web scraping is helpful for acquire data from web and with programming how can we automate all manual task within less amount of time.

I will accomplish this aim by meeting the following objectives:

1. To study the existing approaches of content curation.
2. To present a new technique of content curation with help of web scraping method with automation.
3. Compare the manual outcome of content curation with automate outcome.
4. Determine if automated outcome more relevant to manual outcome.

Web content curation process has extra benefits than saving time. Which makes your task more efficient and easy? In place of human efforts, if you are using an automated system or application, then the user can be capable of curate content more effectively and it's going to additionally helpful to reach out the main goal.

IV. ANALYSIS AND FINDINGS

Workshop analysis, survey questions, and online forums were analyzed using a grounded theory approach -**Corbin and Strauss (2008)**.Grounded principle is observational method that lets you study a particular phenomenon or system and discover new theories which is probably based mostly on the collection and analysis of worldwide information or real world information.

Our research first question: How can be more accurate and relevant Content curation process? After the analysis, we found that content curation is more time consuming task – as content marketer needs to find existing data from 100-1000 of different sources and make relevant or new content. In our surveys we found 80% TIME CONSUMING feedback from content curators and digital marketers. Most of the content curators not getting relevant and updated sources , they said -it's no longer easy to find existing data from hundreds of web sources as we know how world wide web big. This isn't any easy task to managing data Example: Text, Audio, Photos, Videos.
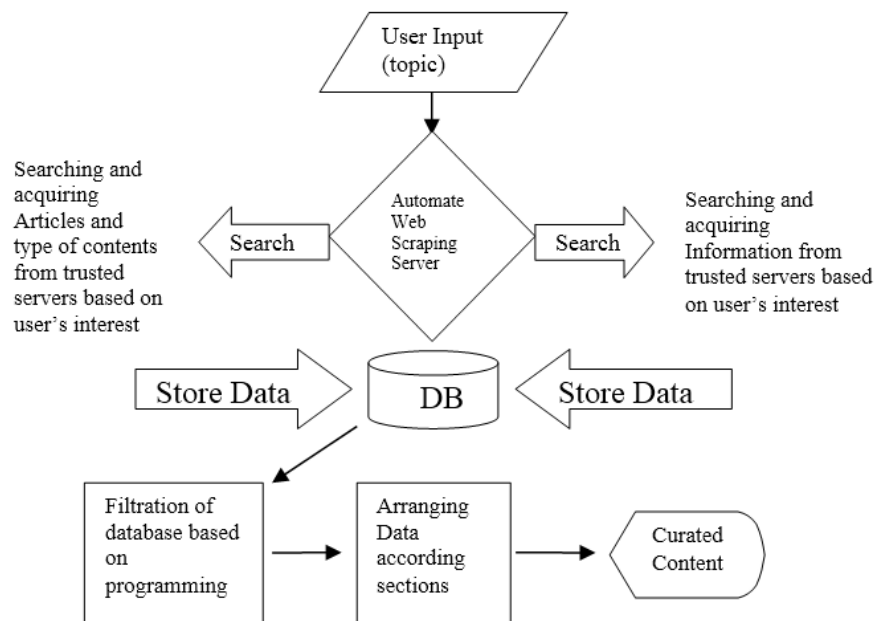

Fig 1: Automated Web Content Curation Flow

V. LIMITATION AND FUTURE SCOPE

This studies method is the overall concept which may be carried out on numerous fields with specific parameters.For example; it is able to be utilized in Journalism, Research paper writing, Digital marketer in which all dealing with different types of data and these fields always trying to search out relevant data. Our method provides them more accurate and relevant data which will go to make their challenge easy and much less time consuming.

In future, The Research outcomes going to be more useful for everyone who are working on data related fields, our search provide more and more accurate data to user for their work, perhaps this going to a portal for data analyst where they find results based on historical and present data. The technique more helpful if user comes with a topic and has not idea about topic, our research technique can gives best results as per their interest.

VI. CONCLUSION

In this paper, the problems behind Content Curation and how today technology is like Web Scraping can restore those problem has been outlined. The different Content Curation techniques and Manual efforts recognized as a trouble and this paper presenting a Web Technique that going to fill this gap.

There are couple of algorithms carried out on this technique. On this technique, via using the use of Web scraping and automation with different internet sources use to help and enhance the results.

In destiny if I get possibility I will make extra researches primarily based totally on same method however in one of a kind fields.From now days as we know Data never going to be old in this world, data is the destiny of latest technology and that's why I am doing this studies and presenting approach which goes to greater precious and really well worth in destiny.

REFERENCES

[1] Neil Beagrie – 2006 - Digital curation for science, http://www.ijdc.net/article/view/6/2

[2] C Zhong, K Sundaravadivelan , S Shah, N Sastry – 2013 - How and why of online content curation., https://ojs.aaai.org/index.php/ICWSM/article/view/14441/14290

[3] J Babin, J Hulland – 2019 - Exploring online consumer curation content: they explained framework for future research https://www.emerald.com/insight/content/doi/10.1108/SJME-07-2019-0053/full/pdf?title=exploring-online-consumer-curation-as-user-generated-content-a-framework-and-agenda-for-future-research-with-implications-for-brand-management

[4] Anand V. Saurkar - 2018 - Web scraping techniques, http://www.ijfrcsce.org/index.php/ijfrcsce/article/view/1529/1529