

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

Automated Aggregation of Internet Content using a RPA Bot

Manishkumar Gami¹, Nidhi Mehta², Parth Jetly³, Dr. Sunita R Patil⁴.

U.G Student, Department of Computer Engineering, KJ Somaiya Institute of Engineering and Information Technology,
Mumbai, Maharashtra, India¹

U.G Student, Department of Computer Engineering, KJ Somaiya Institute of Engineering and Information Technology,
Mumbai, Maharashtra, India²

U.G Student, Department of Computer Engineering, KJ Somaiya Institute of Engineering and Information Technology,
Mumbai, Maharashtra, India³

Vice Principal, Professor, Department of Computer Engineering, KJ Somaiya Institute of Engineering and Information
Technology, Mumbai, Maharashtra, India⁴

ABSTRACT:In today's busy life, precise information is something very difficult to achieve within a few clicks due to infinite content available on the World Wide Web (WWW). This proves the need for content aggregation. Content Aggregation gathers the information by making use of related keywords on a search topic. This is purely an automated task. The bot collects the information based on the keyword and searches for the content as long as the keyword and the user's input are matched. UiPath studio is used for the for implementation of this system. In order to book a flight, user have to navigate the various flights and then select the appropriate result. This paper proposes Robotic Process Automation as a method for performing automated content aggregation of air flight ticket booking through huge information about flights on the internet.

KEYWORDS:Content Aggregation, Robotic Process Automation, Node.js.

I. INTRODUCTION

A bot or a robot is a software or application which can be made to perform malicious tasks or bots can be trained to crawl on the internet looking out for information required by the user. Bots are most popularly used to perform repetitive tasks. Bots can also chat with human and give human like feeling without the user knowing about the fact that it's a machine talking on the other side and not a human. Bot runs a script over the internet to automate the tasks. Growth in technology these days make the average person more and more curious to gain more knowledge on a particular trending topic. This is where content aggregation comes into picture.

The content aggregation is nothing but gathering information on a particular topic using related keywords. There are various number of websites available on the internet providing information about different topics and each website contains plenty of information required by a person to understand the same topic. Content aggregation makes it easy for a person to understand a topic using a small amount of information instead of reading loads of information. Internet contents also known as web contents are huge data available on the internet. This data can be in the form of text, images, videos, animations or audios.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

The proposed system Automated Aggregation of Internet Contents using a RPA bot is nothing but automated software developed to collect content from the internet and then summarizing it as per the user's need through a bot for automated Aggregation of internet contents on a particular topic. The application used for aggregation of internet contents are selected on a air flight booking system. While booking a flight a user checks for prices for a particular source, destination and date of travel from multiple websites available over the internet to select the most affordable and appropriate ticket for himself/herself. This makes the user type into the details again and again for each website.

Using this system, the user will have to give input only once. Thus, saving human time. The systems currently provide output in two ways, two ways one where user intervention is required once and the other where user has to fill in some kind of details again and again to gain some information. The former one includes theoretical topics where the user strives to understand some topic and the later involves working with booking websites like flight booking hotel booking and so on. This system makes use of RPA and node.js as a technology.

Robotic Process Automation: RPA is acronym for robotic process automation. ROBOT – An entity that mimics human action. PROCESS – A series of steps required to perform a task. AUTOMATION – Tasks that don't require human intervention. RPA allows creation of a software using which we can integrate actions which humans performs while working with digital systems. These actions include calculations, maintenance of records, entering data, logging in and logging out, collection of data from file or browser etc. RPA uses human interface to capture and manipulate data. An RPA software robot never ceases to perform its function, is accurate and is cheaper to afford than an employee. An RPA robot can work round the clock without any compromise on the efficiency even after numerous runs in a day.

RPA systems develop series of steps by watching the user perform that task in the application's graphical user interface (GUI), and then perform the automation by repeating those tasks directly in the GUI. RPA provides a platform to create virtual employee. Node.js is an open source JavaScript runtime environment. It executes JavaScript code outside of a browser. The script return in JavaScript are embedded on an HTML webpage. JavaScript unifies web development around a single programming language rather than different languages for server side and client-side programming. Thus, the use of RPA along with Node.js performs the task of automated content aggregation.

II. LITERATURE REVIEW

1. Institute of Exact Sciences and Informatics paper talks about the present experimental results on the proposed methodology which employs ontologies, automatic annotation of metadata, information retrieval and text mining to recommend and aggregate related content, using the relation metadata category as defined by SCORM. It is a computer system prototype which applies the proposed methodology on a sample of learning objects generating results to evaluate its efficacy. The results demonstrate that the proposed method is feasible and effective to produce the expected results.
2. Shan Feng, Zhu Li, Yiling Xu and Jun Sun have developed a novel hash scheme which is based on the deep convolutional network feature maps and Fisher Vector aggregation. Two main contributions in this paper are Fisher Vector aggregation on the features of underlying Gaussian Mixture Model (GMM) and hash generated by direct binarizing the Fisher Vector with component optimization.
3. Munidar P Singh proposes a conceptual framework for enterprise Chabot that employs well-studied abstractions of goals, commitments, and plans. A benefit of using these abstractions is that the complex Chabot can be developed in a flexible manner and as opposed to existing IFTTT frameworks that cause maintenance to be prohibitively resource and cost intensive. According to the study provided in this paper, an engineering Chabot consists of the following components: dialog manager, dialog management, inference engine, knowledgebase and the planner.
4. The paper "Content Aggregation by Platforms: The News Media" shows there are two waves of digitization. In the first wave of digitization, peer-to-peer networks disseminated copyrighted content, sometimes illegally and the second wave of the digitization of content has led to the rise of platforms which aggregate content in a convenient and unified form for consumers. The new practice of digital aggregation has led to both lawsuits and uncertainty over the economic consequences.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

5. Dan He and Douglass S. Parker talk about a personalized content aggregation model which considers the new content posting rate as well as user access rate. Furthermore, this model not only considers the number of missing posts, but also delay time for these missing posts. They developed an optimal resource allocation algorithm, as well as an optimal retrieval scheduling algorithm to minimize the expected delay time of aggregation over multiple data sources.
6. Ruchi Issac, Riya Muni, Kenali Desai state that, to implement RPA, many software platforms are developed and the main purpose of this paper is to provide a detailed analysis of three of the leading RPA platforms namely UiPath Studio, Automation Anywhere and Blue Prism. They provide some graphical data with one simple example of student management system so that we can understand which tool is better amongst the three. Also, in the near future, the scope for UiPath exceeds unimaginable boundaries. This is due to the adaptive algorithms it uses as compared to other two tools.
7. Everest Research Group paper provides detailed information about RPA technology. It examines the RPA technologies for creating a virtual workforce. It assesses ten of the leading technology vendors and compares and contrast their technologies within the feature, implementation and impact matrix framework.
8. Authors Rebecca Dilla and Heidi Jaynes focus specifically on RPA adoption. They have highlighted the benefits and key challenges of RPA. Adoption and growth of RPA related to current market and also for future is well evaluated in this paper. This paper also focuses on what companies should really be considering before finalizing RPA as a solution for their problem statement.
9. Authors Anagnoste and Sorin talk about how robotic process automation is used to solve the major issues faced by many production companies. They propose very simple solution which resolves the production companies issue and also provide cost cutting, better product result with absolutely zero error and also the employee staff who was working in production handling is shifted to other work where human force is required to improve the other product related work. They use technique called Optical Character Recognition (OCR) which is part of RPA.
10. Authors Mary C. Laticy and Leslie P. Willocks have highlighted the importance of Robotic Process Automation technology. They talk about how it can improve the auditing process. The auditing process is very time consuming and complex. Minor error can lead to huge problem and it is a very tedious job to solve that error. They have used automation tools like Excel Macros, IDEA, Python and R for automating different process of auditing like Reconciliations, Analytical Procedures Internal Control Testing Detail Testing (Attribute Match) and reconciliations to develop automated process.

III. SYSTEM DESIGN

In the existing air flight Booking systems, if a user wants to book a flight from source A to destination B then the user has to navigate to different websites and fill in the input again and again. But using Robotic Process Automation solution, the user need not fill the data again and again. This eases up the process of flight booking for the user [11,12,13,14]. The Robotic Process Automation system consists of various stages of design such as Chatbot Creation, Uipath Workflow, Extract Json File, Creation of Node Js and Displaying Output. Figure 1 shows the RPA Automated Aggregation System

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

Design. The system modules are:

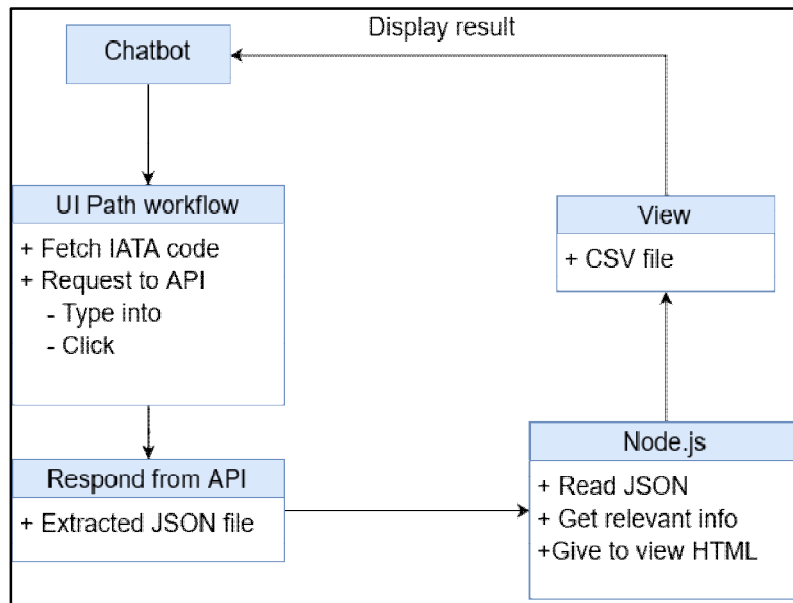


Figure 1 RPA Automated Aggregation System Design

1) Chatbot Creation-

A Chabot is created using the Dialog Flow API. Dialog Flow is a Google-owned developer of human - computer interaction based on natural language processing. Firstly, the user gives input through the Chatbot and the Chabot passes on the data received from the user to the content aggregation bot.

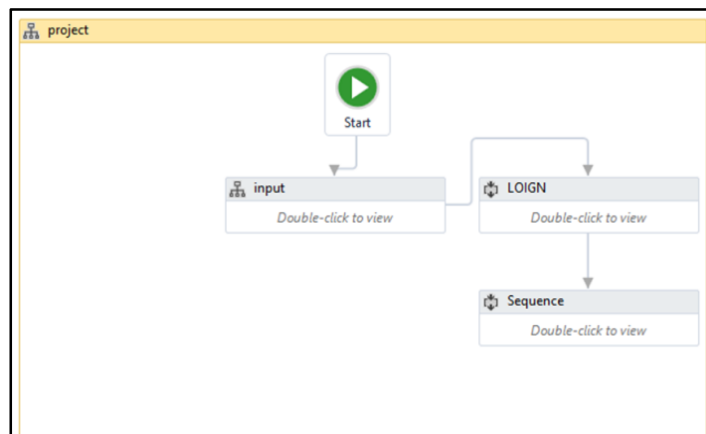


Figure 2. Creation of Input credentials

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

2) UiPath workflow-

The content aggregation bot fetches the IATA code (Mumbai has BOM, Goa has GOI) for the source and the destination provide by the user from the excel file. Then requests for the API and then types into and selects the appropriate value

from the dropdown as per the details provided by the user. The API responds to this request and provides the output in the form of JSON file.

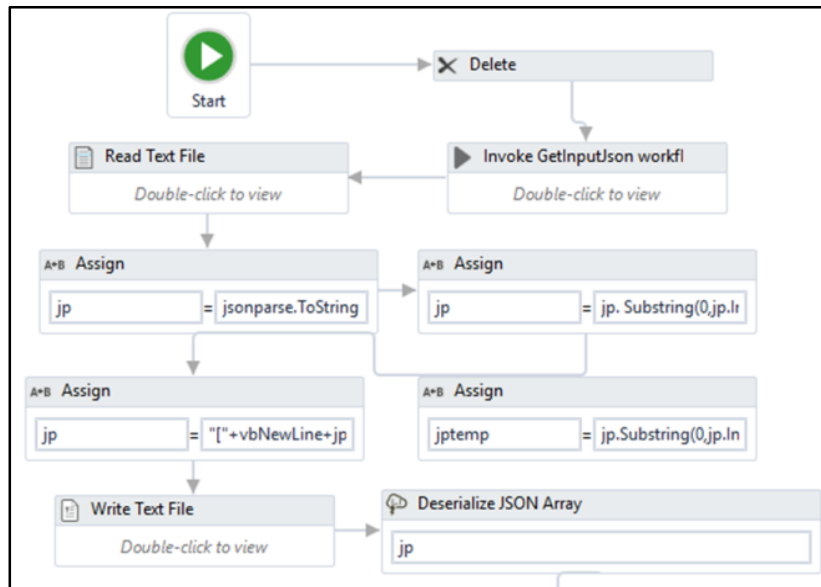


Figure 3. Json file creation.

3) Extract JSON file-

The JSON file available on the output pane of the API is copied for JSON processing.

4) Node.js-

The JSON file thus created in the extraction step is processed using

Node.js cross platform. Node.js firstly reads the JSON file and extracts the required fields. All these fields are viewed with the help of an HTML webpage.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

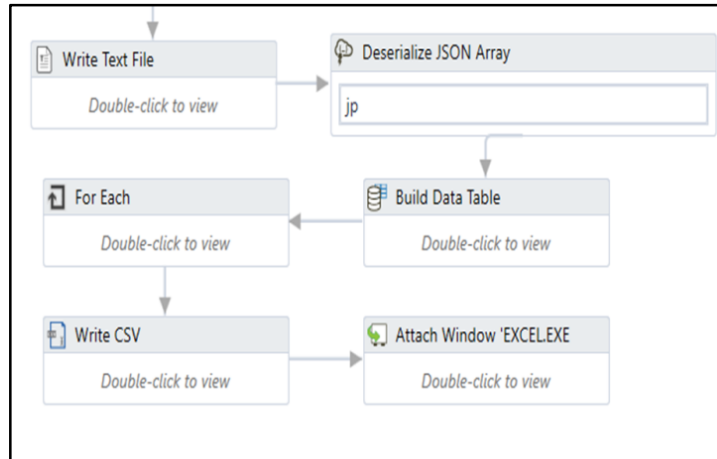


Figure 4. generation of output

5) View

The data is then fetched from the HTML file. Node.js modifies this fetched data and prepares a CSV file which will make the output readable and easily understandable for the user. This CSV file is passed on to the Chabot. Lastly the chatbot displays the expected output to the user.

Table 1. Benefits of automation

Sr. No	BEFORE AUTOMATION	AFTER AUTOMATION
1	No use of chat bot	Making use of chat bot
2	Manual Operations	Automated Operations
3	More time required to get desired result	Less time required to get desired result.
4	More workload	Less workload
5	Environment-centric approach	Centralized backup strategy
6	Giving input multiple times	Giving input single time
7	Less accurate output	More accurate output

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

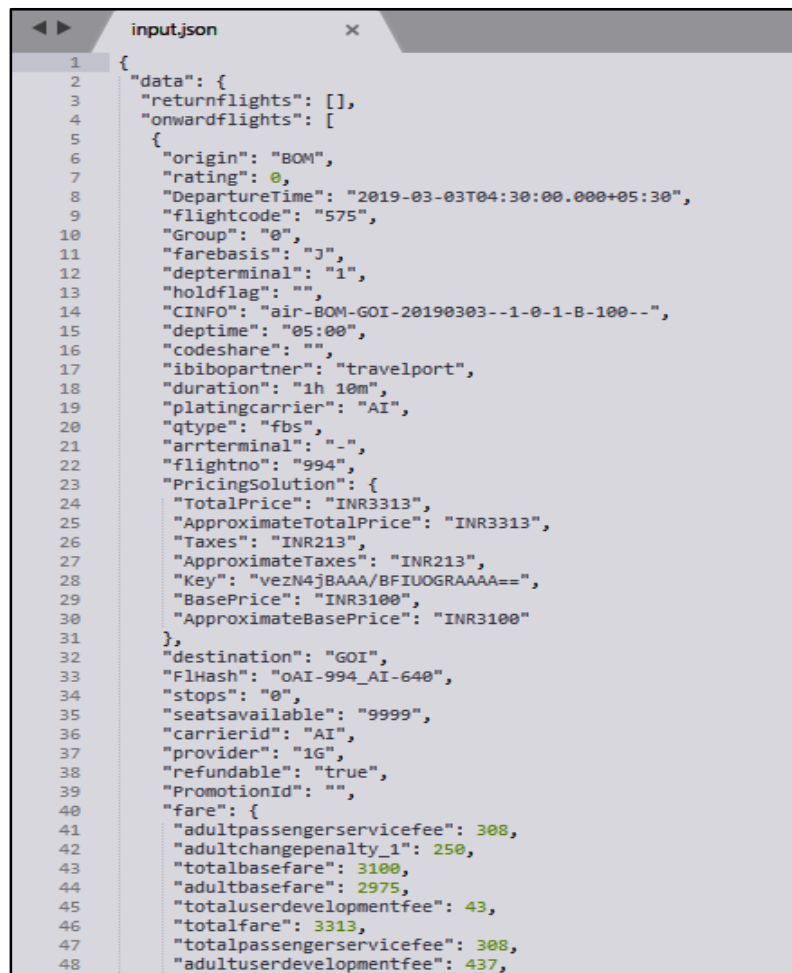
Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

IV. SYSTEM IMPLEMENTATION

The Robotic Process Automation system implementation consists of 3 modules:

- 1) Taking Input Credentials
- 2) Json File Creation
- 3) Generation of Output



```
1 {
2   "data": {
3     "returnflights": [],
4     "onwardflights": [
5       {
6         "origin": "BOM",
7         "rating": 0,
8         "DepartureTime": "2019-03-03T04:30:00.000+05:30",
9         "flightcode": "575",
10        "Group": "0",
11        "farebasis": "J",
12        "depterminal": "1",
13        "holdflag": "",
14        "CINFO": "air-BOM-GOI-20190303--1-0-1-B-100--",
15        "deptime": "05:00",
16        "codeshare": "",
17        "ibibopartner": "travelport",
18        "duration": "1h 10m",
19        "platingcarrier": "AI",
20        "qtype": "fbs",
21        "arrterminal": "-",
22        "flightno": "994",
23        "PricingSolution": {
24          "TotalPrice": "INR3313",
25          "ApproximateTotalPrice": "INR3313",
26          "Taxes": "INR213",
27          "ApproximateTaxes": "INR213",
28          "Key": "ve2N4jBAAA/BFIUOGRAAAA==",
29          "BasePrice": "INR3100",
30          "ApproximateBasePrice": "INR3100"
31        },
32        "destination": "GOI",
33        "FlHash": "oAI-994_AI-640",
34        "stops": "0",
35        "seatsavailable": "9999",
36        "carrierid": "AI",
37        "provider": "IG",
38        "refundable": "true",
39        "PromotionId": "",
40        "fare": {
41          "adultpassengerservicefee": 308,
42          "adultchange penalty_1": 250,
43          "totalbasefare": 3100,
44          "adultbasefare": 2975,
45          "totaluserdevelopmentfee": 43,
46          "totalfare": 3313,
47          "totalpassengerservicefee": 308,
48          "adultuserdevelopmentfee": 437,
```

Figure 5: Snip of JSON file obtained from API

1) Taking input credentials: The user will give input details based on his/her requirements, for example if the user wants to book a flight, he/she has to fill various parameters such as source, destination, date of departure and arrival, class of travel, etc. Once the user has filled in the details, all the user inputs will be taken up by chat bot. The chat bot on receiving the user inputs, logs in into API's of different websites. This is an automated task. Figure 2 shows implementation of module 1 that is Creation of Input credentials.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

```
C:\WINDOWS\system32\cmd.exe - node app.js
C:\Users\wasik\Downloads\Bot>node app.js
Connected on 3100
{ flightcode: '575',
  depttime: '05:00',
  duration: '1h 10m',
  deptterminal: '1',
  origin: 'BOM',
  dest: 'GOI',
  stops: '0',
  airline: 'GoAir',
  seatingclass: 'E',
  totalfare: 3313,
  warning: 'Refundable' }
{ flightcode: '384',
  depttime: '22:05',
  duration: '1h 20m',
  deptterminal: '1',
  origin: 'BOM',
  dest: 'GOI',
  stops: '0',
  airline: 'GoAir',
  seatingclass: 'E',
  totalfare: 1799,
  warning: 'Refundable' }
```

Figure 6: Snip of output obtained using Node.js

2) Json file creation: Once the input information has been filled, API will give the output based on the user input in the form of a json file. Once the json file is created extraction, saving and deserialization of json file is done. This is done using nodes.js.

Figure 3 shows implementation of module 2, That is Json file creation.

3) Generation of Output: The derived json file is converted into text file with the help of write text file activity as shown in the above figure. This text file is further converted into a csv file. This csv file displays the output based on the user input.

Figure 4 shows the implementation of modules 3 that is generation of output.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

Table 1 shows the comparison of before automation and after automation of system aggregate bot. [15,16,17,18]

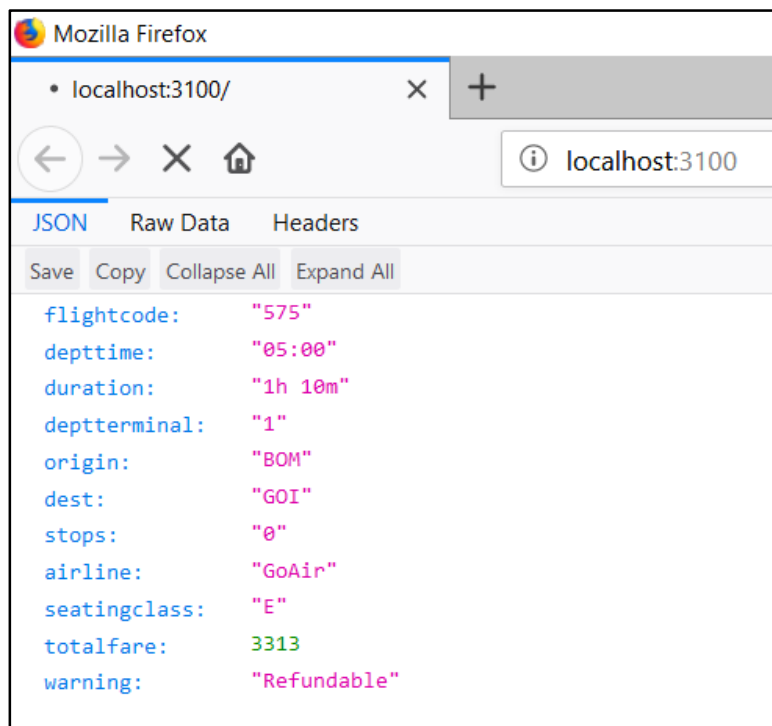


Figure 7: Output as seen on the HTML webpage

V. RESULT

The various output files can be obtained from this Automated Aggregation System are as shown in the below snapshots like:

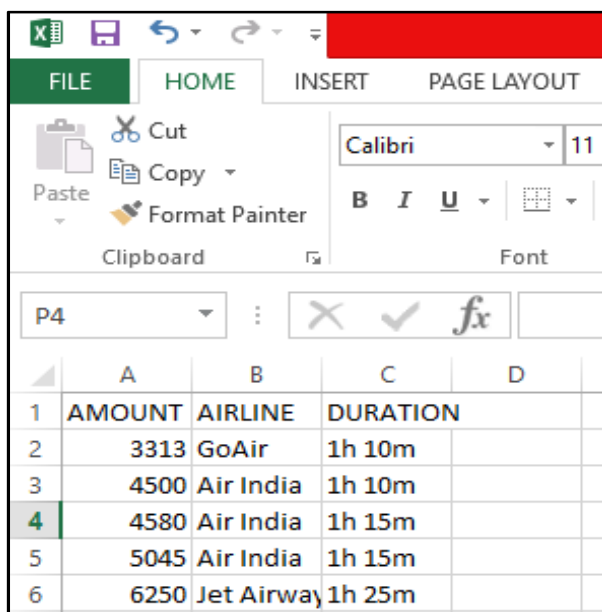
- 1) JSON file - as shown in figure 5
- 2) Processing done using Node.js - as shown in figure 6
- 3) Output as on HTML Webpage - as shown in figure 7
- 4) Sorted CSV File - as shown in figure 8

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019



	A	B	C	D
1	AMOUNT	AIRLINE	DURATION	
2	3313	GoAir	1h 10m	
3	4500	Air India	1h 10m	
4	4580	Air India	1h 15m	
5	5045	Air India	1h 15m	
6	6250	Jet Airway	1h 25m	

Figure 8: Final sorted CSV

[19,20] These results show various benefits of Automated Aggregation bot as compared to existing search engines.

VI. CONCLUSION

The performance evaluation of Automated Aggregation of Internet Content system has a very simple principle and can be easily modified and designed as per user requirement. It can be used by any user to get the result based on the input given such as details regarding to flight booking which consists of various parameters such as source, destination, date of arrival, date of departure, class, etc. These parameters available on internet can consume a lot of time as there are plenty of various sites in which the user has to fill the input individually. This problem can be avoided by using this system where in the user has fill the input details only once and based on the user input the bot will perform the necessary action and display the output to the user. This in turn saves a lot of time of the user and can also increase the efficiency of the system, extending it as per the user requirement.

REFERENCES

- [1] Institute of Exact Sciences and Informatics, Pontifical Catholic University of Minas Gerais, et al. "Automatic Content Recommendation and Aggregation According to SCORM." *Informatics in Education*, vol. 16, no. 2, Sept. 2017, pp. 225–56. *Crossref*, doi:10.15388/infedu.2017.12.
- [2] Shan Feng, Zhu Li, Yiling Xu and Jun Sun Feng "Compact Scalable Hash from Deep Learning Features Aggregation for Content De-Duplication." *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, IEEE, 2017, pp. 1–5. *Crossref*, doi:10.1109/MMSP.2017.8122286..
- [3] Munidar P. Singh, Pankaj R. Telang, Rahul Pandita. "A Conceptual Framework for Engineering Chatbots." *IEEE Internet Computing*, vol. 22, no. 6, Nov. 2018, pp. 54–59. *Crossref*, doi:10.1109/MIC.2018.2877827.
- [4] Chiou, Lesley, and Catherine Tucker. "Content Aggregation by Platforms: The Case of the News Media: JOURNAL OF ECONOMICS & MANAGEMENT STRATEGY." *Journal of Economics & Management Strategy*, Mar. 2017. *Crossref*, doi:10.1111/jems.12207.
- [5] He, Dan, and Douglass S. Parker. "Optimized Retrieval Algorithms for Personalized Content Aggregation." (*Unav*), Aug. 2013. *DataCite*, doi:10.1109/iri.2013.6642482..
- [6] Ruchi Isaac, Riya Muni, Kenali Desai, "Delineated Analysis of Robotic Process Automation Tools." *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAIECC)*, IEEE, 2018, pp. 1–5. *Crossref*, doi:10.1109/ICAIECC.2018.8479511..
- [7] Everest Research Group. "Robotic Process Automation: Technology vendor landscape with fit ma-trix assessment- Technologies for building virtual workforce", December 2016.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

- [8] Capgemini “Robotic Process Automation: Gearing up for greater integration” , 2017 Storyful, 11th August 2017.
- [9] SsuChieh Lin, Lian Hua Shih, Damon Yang, James Lin, Ji Fu Kung, “Apply RPA (Robotic Process Automation) in Semiconductor Smart Manufacturing”, 2018 e-Manufacturing Design Collaboration Symposium (eMDC), 7th September 2018 in IEEE.
- [10] Moffitt, Kevin C., et al. “Robotic Process Automation for Auditing.” *Journal of Emerging Technologies in Accounting*, vol. 15, no. 1, July 2018, pp. 1–10. *Crossref*. doi:10.2308/jeta-10589.
- [11] <https://irpaai.com/what-is-robotic-process-automation/>
- [12] <https://blog.aimultiple.com/rpa/>
- [13] <https://medium.com/@ISHIRInc/robotic-process-automation-predictions-what-is-the-future-of-rpa-989ef8859409>
- [14] <https://www.edureka.co/blog/what-is-robotic-process-automation/>
- [15] [http://www.xeliumtech.com/blog/Why-do-you-need-Robotic-Process-Automation-\(RPA\)](http://www.xeliumtech.com/blog/Why-do-you-need-Robotic-Process-Automation-(RPA))
- [16] <https://www.grantaautomation.co.uk/news/advantages-and-disadvantages-of-robotic-automation/>
- [17] <https://heidicohen.com/content-aggregation-definition/>
- [18] <https://searchcontentmanagement.techtarget.com/definition/content-aggregator>
- [19] <https://www.socialmediatoday.com/content/content-creation-vs-content-aggregation-basics>
- [20] <https://blog.inoreader.com/2018/08/content-aggregation-complete-guide-2018.html>