VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Programme	:	B.Tech	Semester	:	Fall 22-23
Course	:	Robotic Process Automation LAB	Code	:	CSE2023
Faculty	:	Sakthivel V	Slot	:	L29+L30

Date: 14-09-2022

Name: P.Je Sai Kailash

Reg Num: 20BRS1208

Assessment 7

1.) Data Scraping

Aim

Build a workflow using the Data Scraping Wizardthat scrapes blog post titles from the UiPath Blog from multiple pages

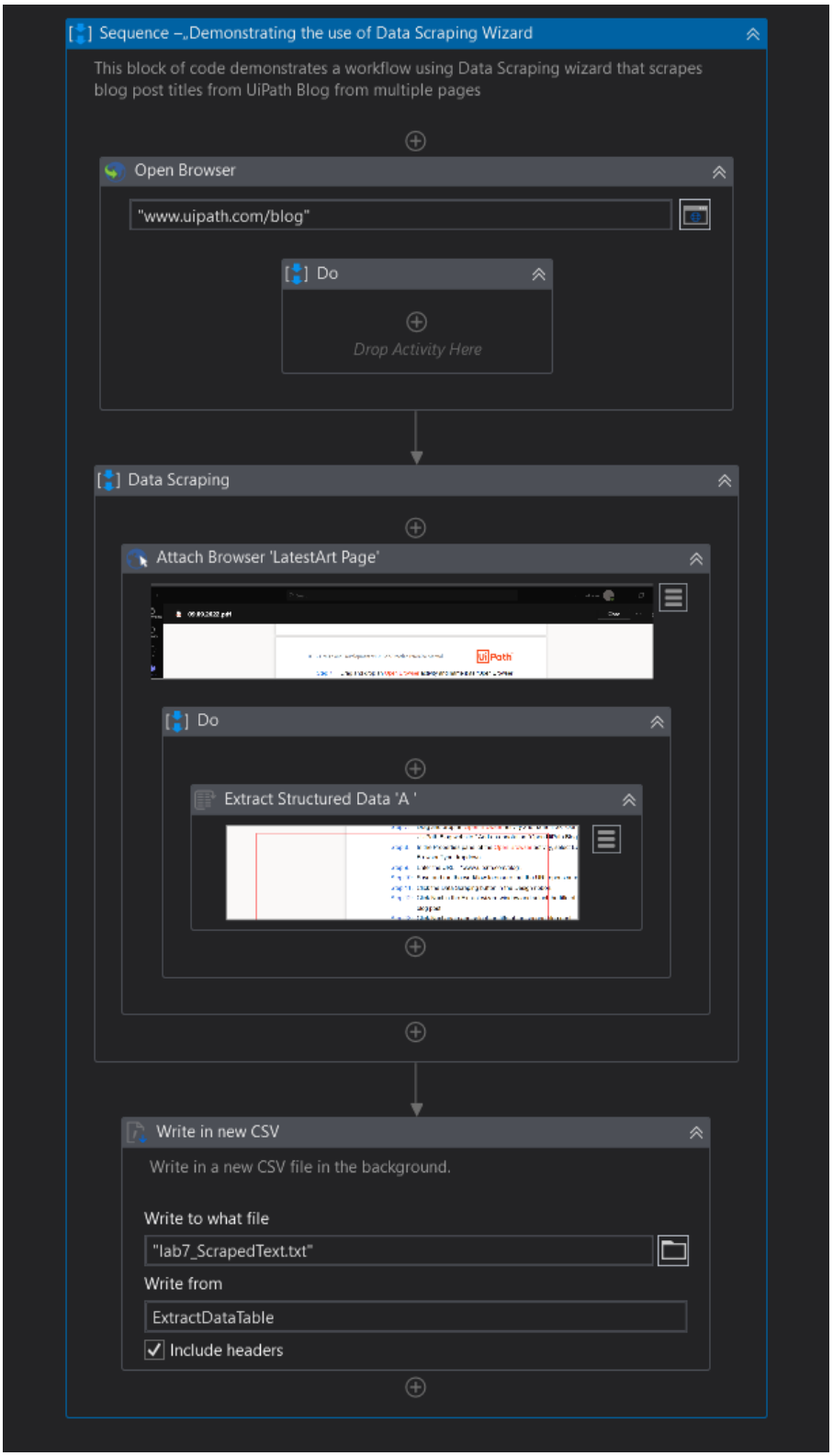
Process Overview

1. Start the process
2. Use an open browser activity and enter url www.uipath.com/blog
3. Use data scraping button in the design ribbon to indicate the first two post titles
4. Rename column1 to blog titles
5. Use the indicate next link window of the data scraping tool to indicate the next link in the search results page
6. Use write CSV activity and save the data in a notepad
7. Stop the process.

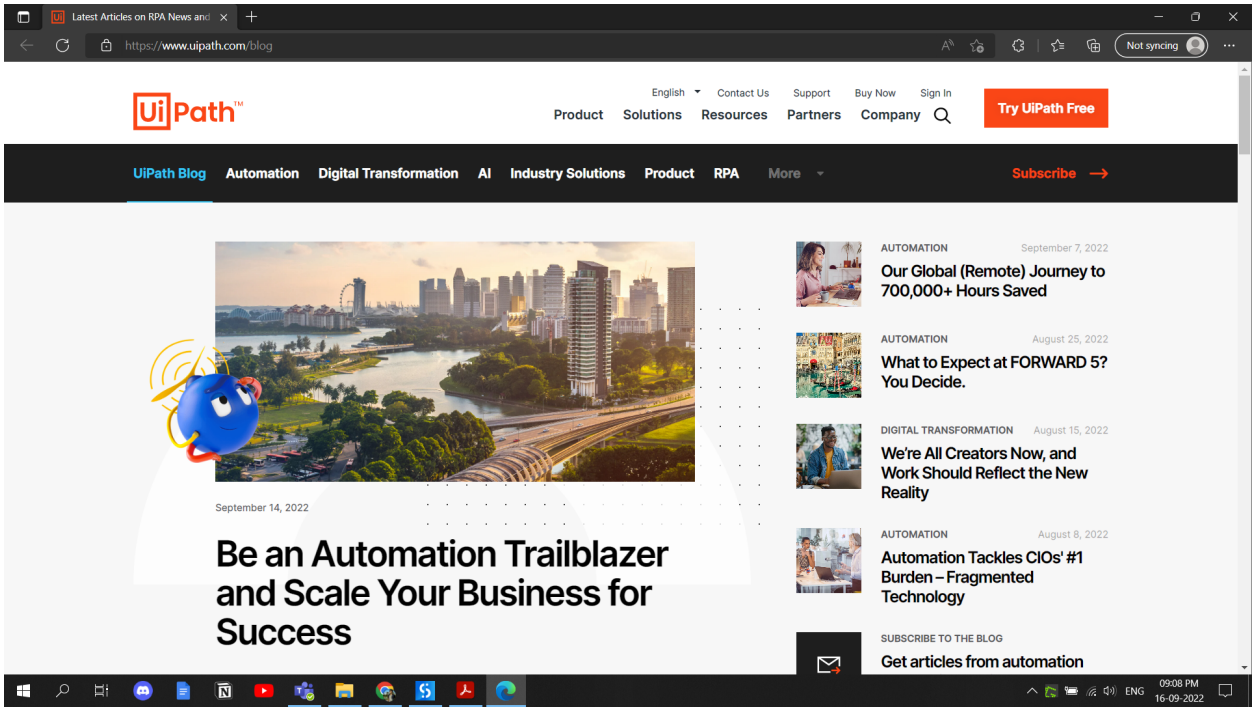
Procedure

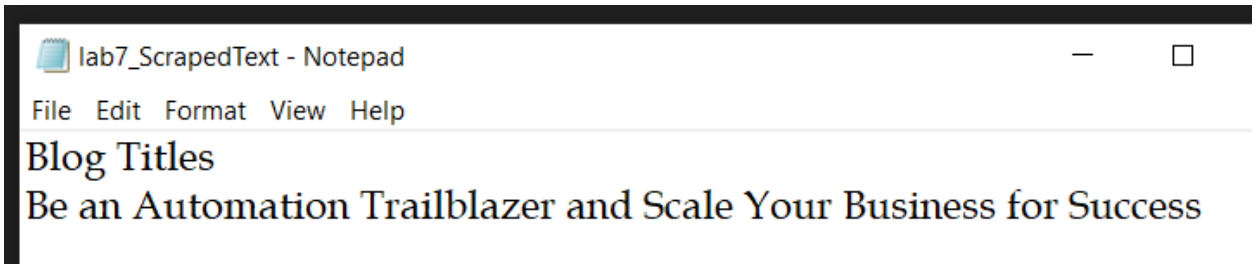
1. Open UiPath.
2. Create a project and start a workflow
3. Add a sequence activity within the workflow
4. Right click on the sequence and select on add annotations, enter the annotations:- “This block of code demonstrates a workflow using Data Scraping wizard that scrapes blog post titles from UiPath Blog from multiple pages”.
5. Now add an open browser activity and in the url section type www.uipath.com/blog
6. select the browser type as edge
7. now click on the data scraping button in the design ribbon
8. click on the next in the extract wizard window and select the first title of a blog post.
9. Now click next and select the title of the second blog post
10. Rename column1 to Blog titles in the data scraping window
11. Now click on finish in the preview window.
12. Now go to the end of the blog page and and locate the load more button
13. Go to the indicate next link window and click yes and indicate it to the load more button
14. Insert a Write csv activity and in the write to what file section enter “ScrapedText.txt” and in the write from section enter the datatable variable where the data is stores.
15. Save, debug and run the file
16. End the process.

Activity


Sequence –„Demonstrating the use of Data Scraping Wizard
This block of code demonstrates a workflow using Data Scraping wizard that scrapes blog post titles from UiPath Blog from multiple pages
Open Browser
"www.uipath.com/blog"
Do
Drop Activity Here
Data Scraping
Attach Browser 'LatestArt Page'
Do
Extract Structured Data 'A '
Write in new CSV
Write in a new CSV file in the background.
Write to what file
"lab7_ScrapedText.txt"
Write from
ExtractDataTable
Include headers

Output


Latest Articles on RPA News and
https://www.uipath.com/blog
Not syncing
UiPath
English
Contact Us
Support
Buy Now
Sign In
Product
Solutions
Resources
Partners
Company
Try UiPath Free
UiPath Blog
Automation
Digital Transformation
AI
Industry Solutions
Product
RPA
More
Subscribe
September 14, 2022
Be an Automation Trailblazer and Scale Your Business for Success
AUTOMATION
September 7, 2022
Our Global (Remote) Journey to 700,000+ Hours Saved
AUTOMATION
August 25, 2022
What to Expect at FORWARD 5? You Decide.
DIGITAL TRANSFORMATION
August 15, 2022
We're All Creators Now, and Work Should Reflect the New Reality
AUTOMATION
August 8, 2022
Automation Tackles CIOs' #1 Burden – Fragmented Technology
SUBSCRIBE TO THE BLOG
Get articles from automation
ENG
09:08 PM
16-09-2022


lab7_ScrapedText - Notepad
File Edit Format View Help
Blog Titles
Be an Automation Trailblazer and Scale Your Business for Success

2.) PDF Extraction

Aim

Build a workflow using the Read PDF Text activity and extract only Email IDs and Phone Numbers from a PDF file and store it in an MS Word file

Process Overview

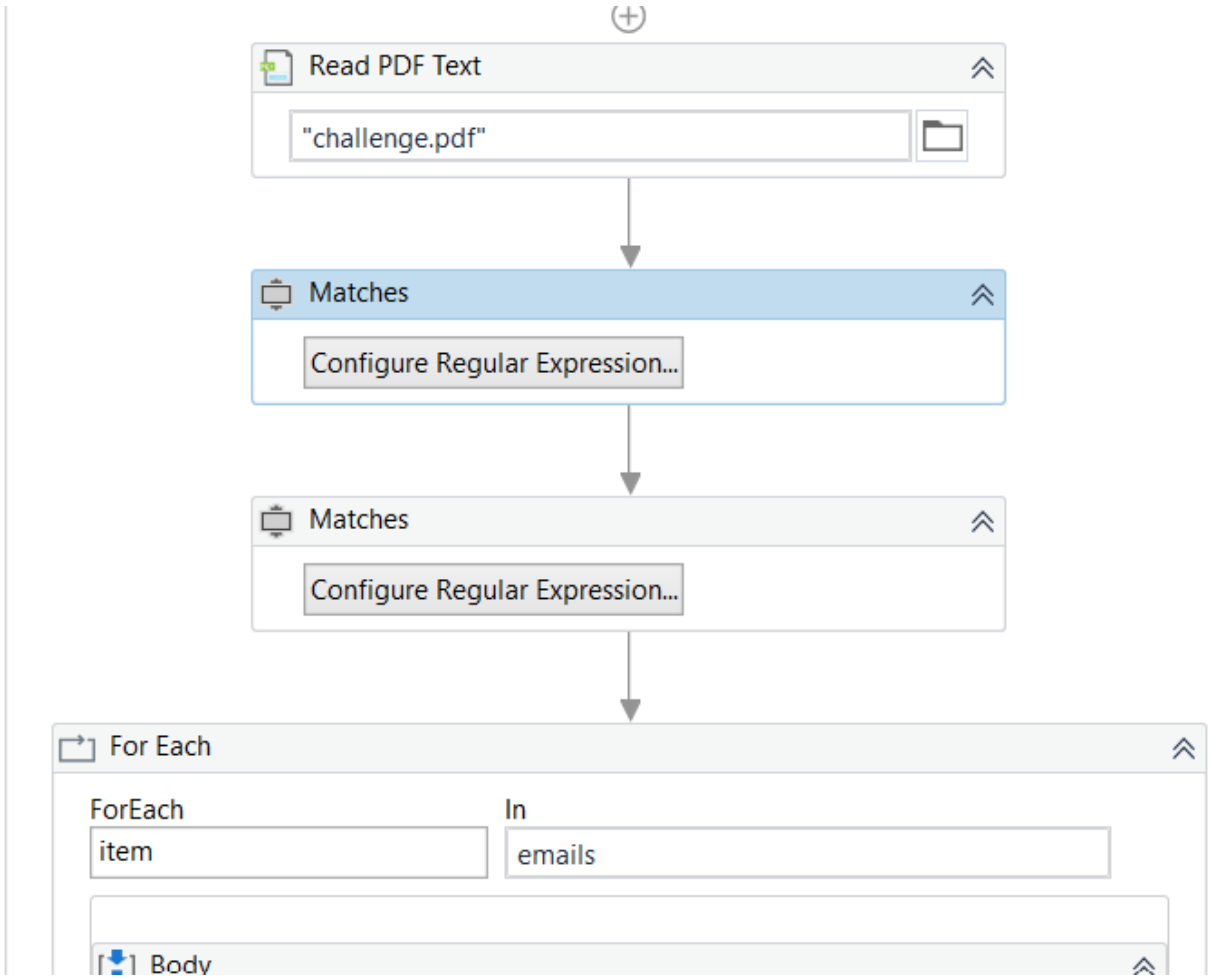
1. Start the process.
2. Use a read PDF text activity to read the contents of the pdf file and store it in a string variable
3. Use a matches activity below the read PDF text activity
a. In RegEx column, select email
4. Use a for each activity to iterate through each email item and store it in an MS word file using a type into activity
5. Use another matches activity below previous matches
a. In RegEx column, select Advanced
b. In Value column, enter the expression : (407)([0-9])
6. Use a for each activity activity to iterate through phone number item and store it in an MS word file using the type into activity
7. Stop the process.

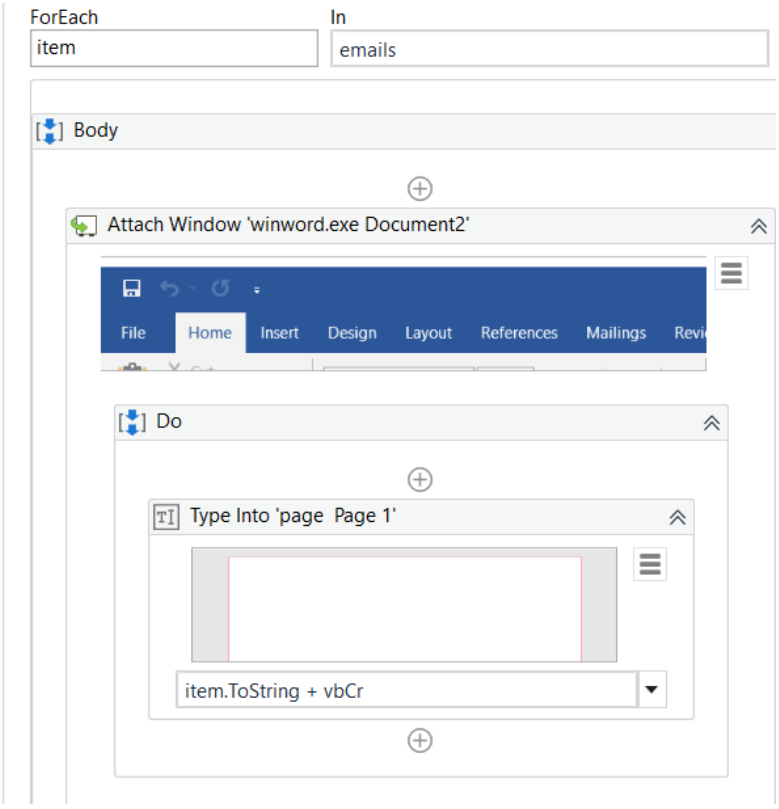
Procedure

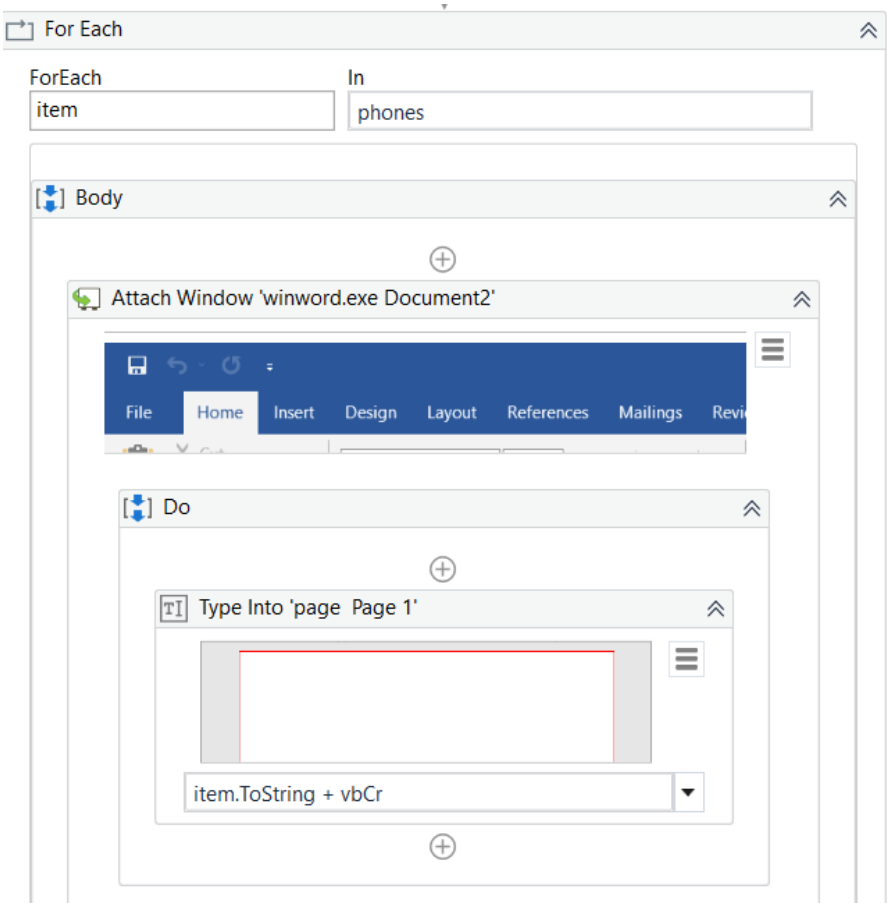
1. Open UiPath
2. Create a project and start a workflow
3. Add a sequence activity within the workflow
4. Now manually open a new MS word
5. Right click on the sequence and add an annotation :- “This block of code demonstrates a workflow using Read PDF Text activity and extract only Email IDs and Phone Number from a PDF file and store in an MS Word file”.
6. Add a read PDF text activity and click the folder and select the pdf file located in the folder.
7. Create a variable named pdfOutput and the variable type should be String and add it into the output property of the properties panel in the read PDF text activity.
8. Insert a matches activity under the read PDF activity and click the configure regular expressions button, in the RegEx column select Email and the quantifiers select Any (0 or more) and click save.
9. The result should be stored in the output variable named as emails.
10. Now add another matches activity below the previous matches activity.
11. Click the configure regular expressions, select the advanced in the RegEx column and in the quantifiers select Any (0 or more)
12. In the value column enter : (407)([0-9])
13. The result should be stored in the output variable named as phones.
14. Now insert a for each activity and in the first box enter “item” and in the second box enter “emails”
15. Insert an attach window activity in the body section and indicate it to the MS word that is opened and confirm the selection
16. In the do section, add a type into activity, click the indicated element on screen and select the editor section of the MS word
17. In the text area enter “item.ToString + vbCr”
18. Now insert another for each activity and in the first box enter “item” and in the second box enter phones.
19. Add a attach window activity in the body section and click on the indicate element on screen and confirm the selection

20. In the do container, add a type into activity and select on the editor section of the MS word
21. In the text area enter “item.ToString + vbCr”
22. Save, debug and run the file
23. End the process.

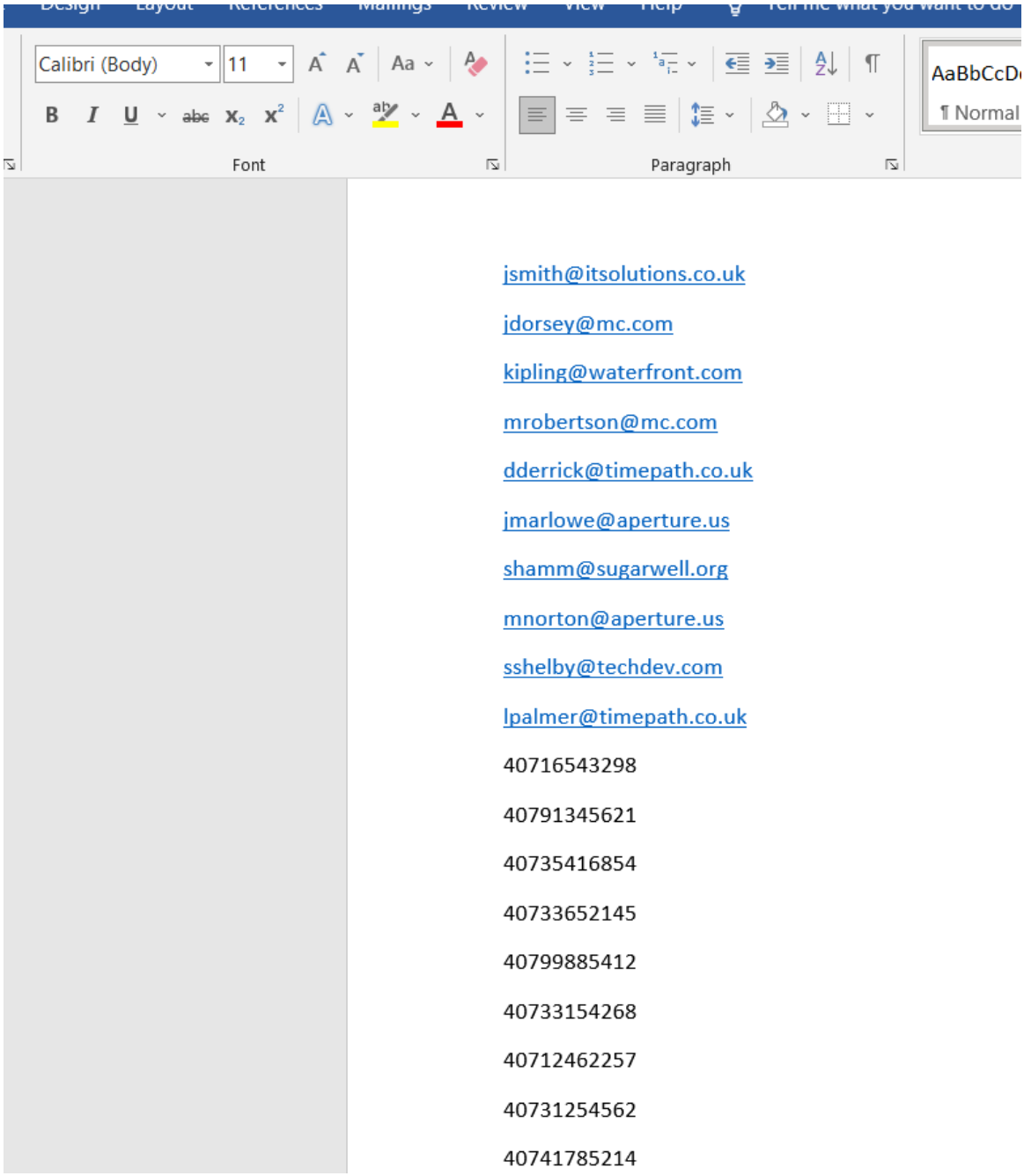
Activity




ForEach
item
In
emails
Body
Attach Window 'winword.exe Document2'
File
Home
Insert
Design
Layout
References
Mailings
Do
Type Into 'page Page 1'
item.ToString + vbCr


For Each
ForEach
item
In
phones
Body
Attach Window 'winword.exe Document2'
File
Home
Insert
Design
Layout
References
Mailings
Do
Type Into 'page Page 1'
item.ToString + vbCr

Output



Result

Hence, we have successfully performed data scraping and also extracted data from a pdf and typed it into a word document.