Development of VBA Tool for Document Term Search

Lemuel Raúl Rios Santiago

NASA's Kennedy Space Center

Major: Industrial Engineering

Systems Development, Data Mining and Knowledge Discovery Program
Fall Session

Date: 10 11 2020

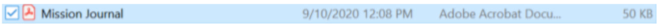
I. Abstract

Employees throughout different agencies such as NASA, have identified that the search of determined terms/words through documents, consume substantial research time of such. These types of searches are substantially limited towards one word in a one document identification; forward one, these usual types of searches lack efficiency & optimization through research aspects of work. Consequently, this reflects in the decrease productivity during work hours etc. The application of *VBA* (*Visual Basic for Applications*) is the programming language of Excel, which was conducted for the development of optimized tool for document term search. The project enables the search of single & multiple word/term search through single format documents for paragraph data extraction.

II. Materials & Methods

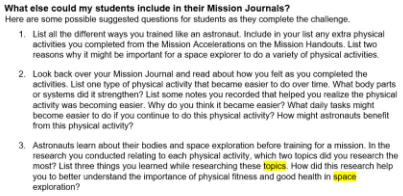
How to Guide: Development for Single & Multiple Word Search through Single File:

Figure 1. **Open PDF Document of Preference**.



Mission Journal 9/10/2020 12:08 PM Adobe Acrobat Docu... 50 KB

Figure 2. **Identify term(s) to search in VBA and leave opened the PDF software.**

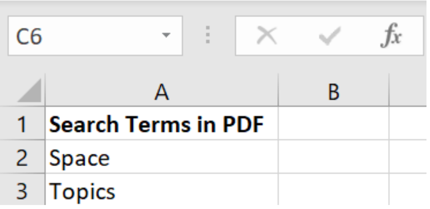


What else could my students include in their Mission Journals?
Here are some possible suggested questions for students as they complete the challenge.

1. List all the different ways you trained like an astronaut. Include in your list any extra physical activities you completed from the Mission Accelerations on the Mission Handouts. List two reasons why it might be important for a space explorer to do a variety of physical activities.
2. Look back over your Mission Journal and read about how you felt as you completed the activities. List one type of physical activity that became easier to do over time. What body parts or systems did it strengthen? List some notes you recorded that helped you realize the physical activity was becoming easier. Why do you think it became easier? What daily tasks might become easier to do if you continue to do this physical activity? How might astronauts benefit from this physical activity?
3. Astronauts learn about their bodies and space exploration before training for a mission. In the research you conducted relating to each physical activity, which two topics did you research the most? List three things you learned while researching these topics. How did this research help you to better understand the importance of physical fitness and good health in space exploration?

Figure 3. **Open Excel.**

Figure 4. **Copy and Paste term(s) in Excel and create a column designated for the search of such in MS Word through VBA.**



C6

	A	B
1	Search Terms in PDF	
2	Space	
3	Topics	

Figure 5. **Enable Developer Tab.**

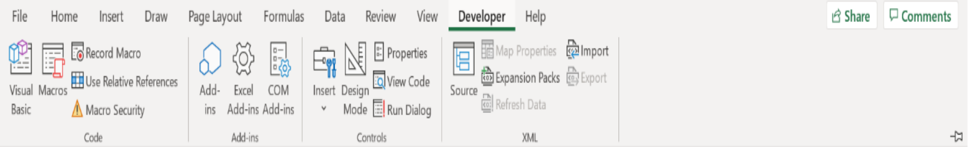
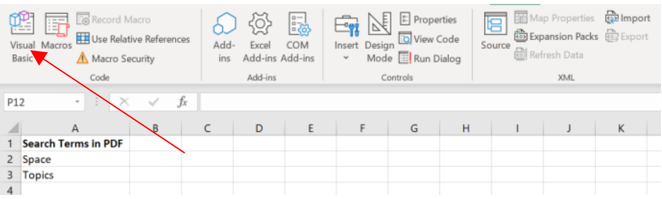


Figure 6. **Open Visual Basic Tab.**



	A	B	C	D	E	F	G	H	I	J	K
1	Search Terms in PDF										
2	Space										
3	Topics										
4											

Figure 7. **Click on Module Tab.**

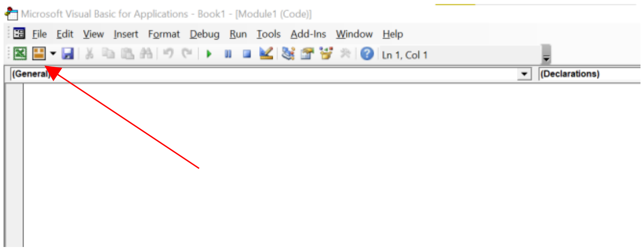
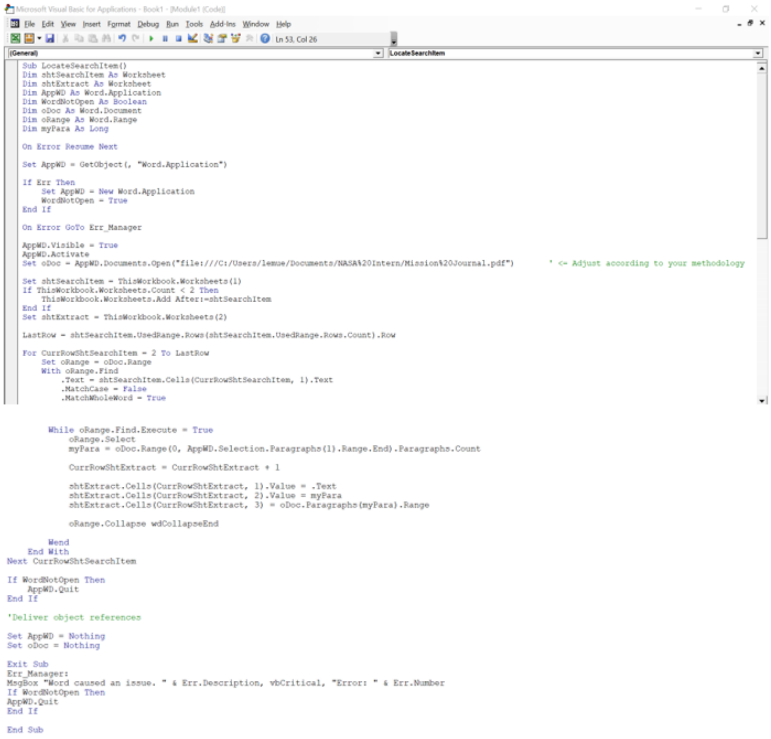
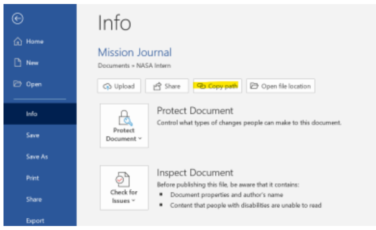


Figure 8. **Insert developed code**.



```vba
Sub LocateSearchItem()
Dim shtSearchItem As Worksheet
Dim shtExtract As Worksheet
Dim AppWD As Word.Application
Dim WordNotOpen As Boolean
Dim oDoc As Word.Document
Dim oRange As Word.Range
Dim myPara As Long

On Error Resume Next

Set AppWD = GetObject(, "Word.Application")

If Err Then
    Set AppWD = New Word.Application
    WordNotOpen = True
End If

On Error GoTo Err_Manager

AppWD.Visible = True
AppWD.Activate
Set oDoc = AppWD.Documents.Open("file:///C:/Users/lemue/Documents/NASA%20Intern/Mission%20Journal.pdf")       ' <= Adjust according to your methodology

Set shtSearchItem = ThisWorkbook.Worksheets(1)
If ThisWorkbook.Worksheets.Count < 2 Then
    ThisWorkbook.Worksheets.Add After:=shtSearchItem
End If
Set shtExtract = ThisWorkbook.Worksheets(2)

LastRow = shtSearchItem.UsedRange.Rows(shtSearchItem.UsedRange.Rows.Count).Row

For CurrRowShtSearchItem = 2 To LastRow
    Set oRange = oDoc.Range
    With oRange.Find
        .Text = shtSearchItem.Cells(CurrRowShtSearchItem, 1).Text
        .MatchCase = False
        .MatchWholeWord = True


        While oRange.Find.Execute = True
            oRange.Select
            myPara = oDoc.Range(0, AppWD.Selection.Paragraphs(1).Range.End).Paragraphs.Count

            CurrRowShtExtract = CurrRowShtExtract + 1

            shtExtract.Cells(CurrRowShtExtract, 1).Value = .Text
            shtExtract.Cells(CurrRowShtExtract, 2).Value = myPara
            shtExtract.Cells(CurrRowShtExtract, 3) = oDoc.Paragraphs(myPara).Range

            oRange.Collapse wdCollapseEnd

        Wend
    End With
Next CurrRowShtSearchItem

If WordNotOpen Then
    AppWD.Quit
End If

'Deliver object references

Set AppWD = Nothing
Set oDoc = Nothing

Exit Sub
Err_Manager:
MsgBox "Word caused an issue. " & Err.Description, vbCritical, "Error: " & Err.Number
If WordNotOpen Then
AppWD.Quit
End If

End Sub
```

Figure 9. **Copy Path in section "AppWD.Documents.Open" of code (File, Info, Copy Path, Paste in Code) (In MS Word); (Select PDF Doc in files, Open with, Internet Explorer, Copy link from search bar (AppWD.Documents.Open("file:///C:/Users/lemue/Documents/NASA%20Intern/Mission%20Journal.pdf")), Paste in Code) (In PDF).**



Mission Journal

Documents » NASA Intern

Protect Document

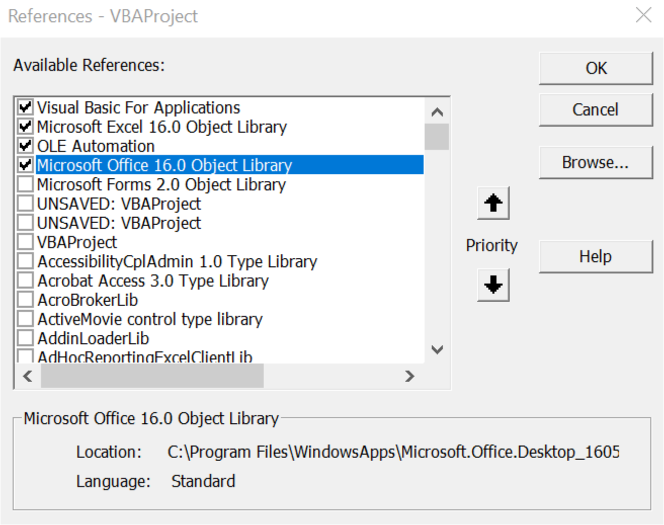
Control what types of changes people can make to this document.

Inspect Document

Before publishing this file, be aware that it contains:

- Document properties and author's name
- Content that people with disabilities are unable to read

Figure 10. **Enable MS Office 16.0 OL that will allow code to run effectively.**



References - VBAProject

Available References:

☑ Visual Basic For Applications
☑ Microsoft Excel 16.0 Object Library
☑ OLE Automation
☑ Microsoft Office 16.0 Object Library
□ Microsoft Forms 2.0 Object Library
□ UNSAVED: VBAProject
□ UNSAVED: VBAProject
□ VBAProject
□ AccessibilityCplAdmin 1.0 Type Library
□ Acrobat Access 3.0 Type Library
□ AcroBrokerLib
□ ActiveMovie control type library
□ AddinLoaderLib
□ AdHocReportingExcelClientLib

Microsoft Office 16.0 Object Library

Location: C:\Program Files\WindowsApps\Microsoft.Office.Desktop_1605

Language: Standard

Figure 11. **Go to Insert and select Button (Form Control).**

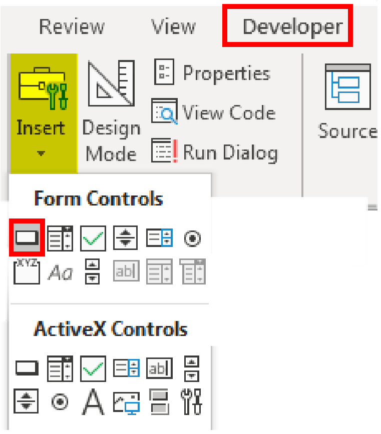
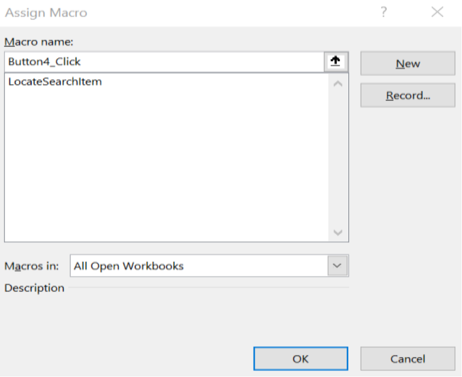


Figure 12. **Select Locate Search Item to enable button that will function to run the VBA code from Excel.**

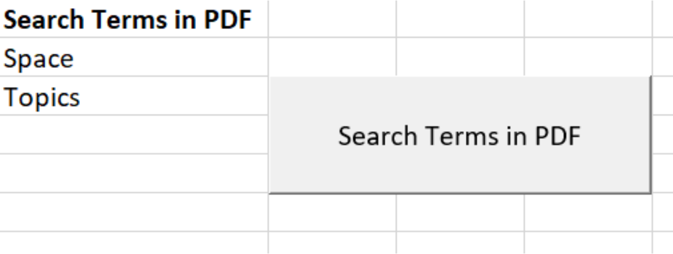


Macro name:
Button4_Click
LocateSearchItem

Macros in: All Open Workbooks

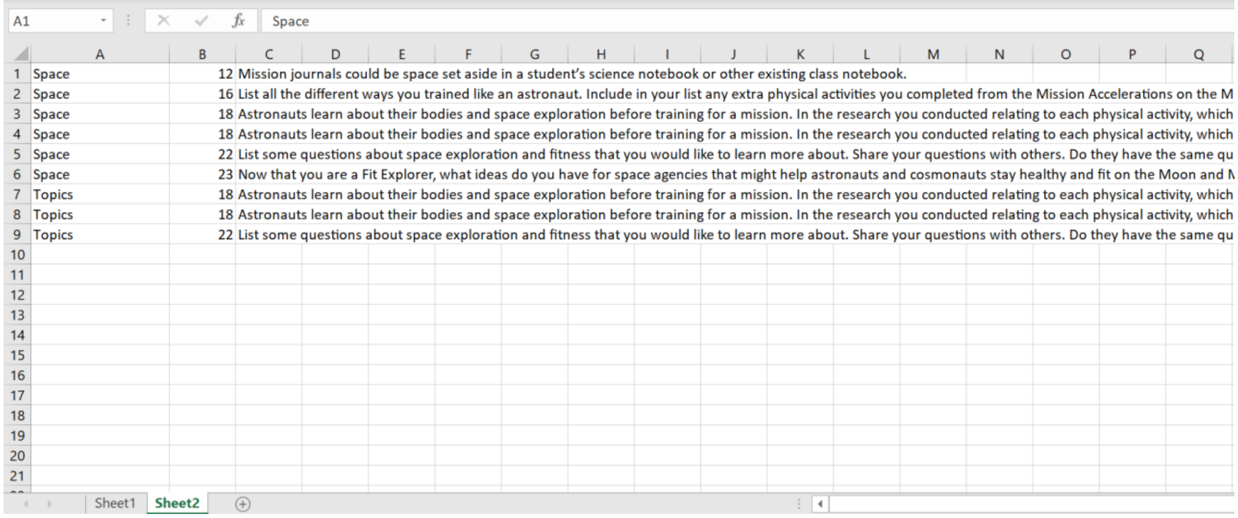
Description

Figure 13. **Change name of Button to Search Terms in PDF for optimized definition of function and then click on such to run the VBA Code.**



Search Terms in PDF
Space
Topics

Figure 14. **The VBA Code will convert the PDF document into an MS Word Document; automatically it will identify the selected terms in Excel and display such through Sheet 2 within the sentences such appeared.**



A1 | Space

	A	B	C
1	Space	12	Mission journals could be space set aside in a student's science notebook or other existing class notebook.
2	Space	16	List all the different ways you trained like an astronaut. Include in your list any extra physical activities you completed from the Mission Accelerations on the M
3	Space	18	Astronauts learn about their bodies and space exploration before training for a mission. In the research you conducted relating to each physical activity, which
4	Space	18	Astronauts learn about their bodies and space exploration before training for a mission. In the research you conducted relating to each physical activity, which
5	Space	22	List some questions about space exploration and fitness that you would like to learn more about. Share your questions with others. Do they have the same qu
6	Space	23	Now that you are a Fit Explorer, what ideas do you have for space agencies that might help astronauts and cosmonauts stay healthy and fit on the Moon and M
7	Topics	18	Astronauts learn about their bodies and space exploration before training for a mission. In the research you conducted relating to each physical activity, which
8	Topics	18	Astronauts learn about their bodies and space exploration before training for a mission. In the research you conducted relating to each physical activity, which
9	Topics	22	List some questions about space exploration and fitness that you would like to learn more about. Share your questions with others. Do they have the same qu
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			

III. Results & Discussion

VBA Code Results & Explanation for Finding Terms:

```
Sub LocateSearchItem()
    ‘Module name
    Dim shtSearchItem As Worksheet
    ‘Enables search of Items in VBA through Word
    Dim shtExtract  As Worksheet
    ‘Extracts Data from Word To Worksheet
    Dim AppWD       As Word.Application
    ‘Recognizes Office 365/Word Application
    Dim WordNotOpen As Boolean
    ‘Indicates If MS Word Is Not open
    Dim oDoc        As Word.Document
    ‘Indicates type of document is an MS Word document type
    Dim oRange      As Word.Range
    ‘This range would refer to first word/term in document
    Dim myPara      As Long
    ‘Ensures paragraph collection in document

    On Error Resume Next
    ‘Resumes towards next section after error
    Set AppWD = GetObject(, "Word.Application")
    ‘Determines MS Word for code use
    If Err Then
        Set AppWD = New Word.Application
        ‘Enables new MS Word application after error
        WordNotOpen = TRUE
        ‘Indicates that Word Is not open
    End If
    Error handling in Case MS Word Is already open
    On Error GoTo Err_Manager
    ‘If an Error Is detected, a message pertaining the Error will be displayed below at “Err_Manager:”
```

```
AppWD.Visible = TRUE
‘Enables MS Word to be visible
AppWD.Activate
‘Activates MS Word
Set oDoc = AppWD.Documents.Open("file:///C:/Users/lemue/Documents/NASA%20Intern/Mission%20Journal.pdf")        ' <= Adjust according to your methodology

Set shtSearchItem = ThisWorkbook.Worksheets(1)
‘Worksheet 1 will enable the process of search of Items
If ThisWorkbook.Worksheets.Count < 2 Then
    ThisWorkbook.Worksheets.Add After:=shtSearchItem
    ‘If sheet 2 Is not enabled, such will automatically be enabled when code Is Set To run; thus, this will process the searched items in sheet 2
End If
Set shtExtract = ThisWorkbook.Worksheets(2)

LastRow = shtSearchItem.UsedRange.Rows(shtSearchItem.UsedRange.Rows.Count).Row
‘This determines as used range process will be used in locating the total count of rows in the worksheet
For CurrRowShtSearchItem = 2 To LastRow
    ‘Initial code looping process
    Set oRange = oDoc.Range
    ‘Set range towards document that Is being used
    With oRange.Find
        .Text = shtSearchItem.Cells(CurrRowShtSearchItem, 1).Text
        ‘Compares Text in Sheet Search Item and assigns Text obtained from oRange
        .MatchCase = FALSE
        ‘Set to FALSE in case wrong letter or format Is typed for search
        .MatchWholeWord = TRUE
        ‘Set to TRUE if it equals to the same word in the doc that is being typed for search capabilities
        While oRange.Find.Execute = TRUE
            ‘Find Text capability Is enabled and is true
```

```
          oRange.Select
          'Range selection
          myPara = oDoc.Range(0, AppWD.Selection.Paragraphs(1).Range.End).Paragraphs.Count
          'Obtain paragraph number in which the search item is located

          CurrRowShtExtract = CurrRowShtExtract + 1
          'Location where data will be established

          shtExtract.Cells(CurrRowShtExtract, 1).Value = .Text
          'Text that will be discovered will be placed here
          shtExtract.Cells(CurrRowShtExtract, 2).Value = myPara
          'Obtained paragraph data will be placed here
          shtExtract.Cells(CurrRowShtExtract, 3) = oDoc.Paragraphs(myPara).Range
          'Extraction of complete paragraph will be placed here
          oRange.Collapse wdCollapseEnd
          'Ensures looping process starts With a While (oRange) and ends With a Wend; furthermore, With ends With End With
        Wend
      End With
    Next CurrRowShtSearchItem
    'For CurrRowShtSearchItem ends With Next
    If WordNotOpen Then
      AppWD.Quit
      'If Word Is Not open, such will quit
    End If
    'Deliver Object references
    Set AppWD = Nothing
    Set oDoc = Nothing
    'Release of memory (Permits Macro To run better in next occasion/run)

    Exit Sub
    Err_Manager:
```

```
    MsgBox "Word caused an issue. " & Err.Description, vbCritical, "Error: " & Err.Number
    'Display of Error message
    If WordNotOpen Then
        AppWD.Quit
        'Word file that is open, will close
    End If
End Sub
```

Code Discussion: The code will run from a designated command button of Excel, to detect determined words from Excel cells to PDF & MS Word documents. After identifying the determined words in such documents, the code will proceed to extract the word data through paragraph structure, from the sections that such words appear; thus, if the document is a PDF file, the code will convert such to an MS Word file for an optimized data extraction. The data extracted will appear in a second sheet that the code will automatically enable if such is not available; thus, the data will be divided in two columns. The first column will have the determined words that were searched, and the second column will have the paragraphs where the words appeared. This code has the capability to do single & multiple word/term search through single format documents for paragraph data extraction.

IV. Conclusions

The VBA tool for document term search, aims to help the user search for words through single & multiple word capability in single & multiple document files. Furthermore, allow the user to do such search from an optimized & simple VBA GUI (Graphical User Interface) for visual ergonomic methodologies. The project currently meets the project objectives of executing single & multiple word search through single file documents. The project end objectives will be assessed through the end of the internship fall term, such as the capability of executing single & multiple term search through multiple documents in a single code run (Loop code for folder & subfolders search integration). Additionally, the development of compatible code & design for GUI integration of program algorithm will be assessed for user friendly experience.

Acknowledgments

I would like to express my deepest gratitude to Dr. Shaykhian whose enormous project mentoring, support and insightful comments were invaluable during my project. I am also in debt to Tristian Running Crane, Janelisse Morales and Makayla Renfro whose opinions and information have helped me very much throughout the development of this internship experience. I would also like to express my gratitude to my family for their moral support and warm encouragements. Finally, I gratefully appreciate the financial support of USRA Stipend that made it possible to complete my internship.

References

"Excel Insert Button: How to Insert a Button in Excel with Examples," *EDUCBA* Available: https://www.educba.com/excel-insert-button/.

Excel Off The Grid, "VBA code to loop through files in a folder (and sub folders)," *Excel Off The Grid* Available: https://exceloffthegrid.com/vba-code-loop-files-folder-sub-folders/.

Ihor, "Microsoft Excel Logo," *1000 Logos The Famous Brands and Company Logos in the World Microsoft Excel Logo Comments* Available: https://1000logos.net/microsoft-excel-logo/.

"Loop in Excel VBA," *Excel Tutorial* Available: https://www.excel-easy.com/vba/loop.html.

Newman, C., "How To Build A Modern-Looking VBA Userform," *The Spreadsheet Guru* Available: https://www.thespreadsheetguru.com/blog/build-modern-vba-userforms.

Takyar, A. D. K., "How to Loop through SubFolders," *Free Excel Training Videos* Available: https://www.exceltrainingvideos.com/how-to-loop-through-subfolders/.

Takyar, A. D. K., "Extract Specific Data from MS Word into Excel with VBA," *Free Excel Training Videos* Available: https://www.exceltrainingvideos.com/extract-specific-data-from-ms-word-into-excel-with-vba/.

Tommy ZTommy Z 56955 gold badges99 silver badges1717 bronze badges, ForceMagicForceMagic 5, and ivan_pozdeevivan_pozdeev 26.9k1010 gold badges7777 silver badges126126 bronze badges, "User defined type not defined- controlling Word from Excel," *Stack Overflow* Available: https://stackoverflow.com/questions/11728899/user-defined-type-not-defined-controlling-word-from-excel/17412117.

"What is a Mission Journal?"