Methods/R Code/ Results/Outputs

To explore the PDF and Images in Linux I downloaded some packages in order to help me manipulate the data. I decided to use CentOS (7) since I was most familiar with this system and made sure I had the most up to date version. Once the packages were downloaded, I then uploaded the Kashimr Wallflower pdf and ensured it was in y directory and ready to use.

```
[vagrant@c7401 ~]$ wget http://archive.org/download/WildFlowersOfKashmir/KashmirWildflowers.pdf
--2019-04-13 17:25:14--  http://archive.org/download/WildFlowersOfKashmir/KashmirWildflowers.pdf
Resolving archive.org (archive.org)... 207.241.224.2
Connecting to archive.org (archive.org)|207.241.224.2|:80... connected.
HTTP request sent, awaiting response... 302 Found
Location: http://ia801602.us.archive.org/21/items/WildFlowersOfKashmir/KashmirWildflowers.pdf [following]
--2019-04-13 17:25:15--  http://ia801602.us.archive.org/21/items/WildFlowersOfKashmir/KashmirWildflowers.pdf
Resolving ia801602.us.archive.org (ia801602.us.archive.org)... 207.241.228.122
Connecting to ia801602.us.archive.org (ia801602.us.archive.org)|207.241.228.122|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3945597 (3.8M) [application/pdf]
Saving to: 'KashmirWildflowers.pdf'

100%[==========================================================================================================>] 3,945,597   3.23MB/s   in 1.2s

2019-04-13 17:25:16 (3.23 MB/s) - 'KashmirWildflowers.pdf' saved [3945597/3945597]
```

```
[vagrant@c7401 ~]$ ls
anaconda3                 infernotxt.txt                          KashmirWildflowers.pdf  QueryResults.csv
inferno00dant_2_djvu.txt  infernotxxt.txt                         pdftk-2.02-             rstudio-server-rhel-1.1.463-i686.rpm
inferno00dant_2_djvu.txt.1  Inverted_Index-Starter-Shell-Copy2.ipynb  pdftk-2.02-.1           shakespeare.txt
inferno00dant_2_djvu.txt.2  Inverted_Index-Starter-Shell.ipynb       QueryResults3.csv       spark-wordcount.ipynb
[vagrant@c7401 ~]$
```

Next, I converted the pdf to a text file kwf.txt, and made sure it was in the directory for use. I then looked at the file with the less command and then made sure that I could use the grep command to look at the China color example and was successful.

```
[vagrant@c7401 ~]$ ls
anaconda3                 inferno00dant_2_djvu.txt.2              KashmirWildflowers.dsc  pdftk-2.02-.1                    shakespeare.txt
epel-release-7-11.noarch.rpm  infernotxt.txt                      KashmirWildflowers.pdf  pdftk-2.02-1.el6.x86_64.rpm      spark-wordcount.ipynb
index.html                infernotxxt.txt                         kwf.txt                 QueryResults3.csv
inferno00dant_2_djvu.txt  Inverted_Index-Starter-Shell-Copy2.ipynb  KWF.txt                 QueryResults.csv
inferno00dant_2_djvu.txt.1  Inverted_Index-Starter-Shell.ipynb     pdftk-2.02-             rstudio-server-rhel-1.1.463-i686.rpm
```

```
INDIAN AGRWUfJ'!'URAL

RESEARCH INSTITUTE.

Nuv

5 7 ~27
L.A.A 1.<1.

CHP NLK-a·3 I.A.H..L-I0 5"55-~1 S,UQO

DELHI

^L^L^LWILD FLOWERS
OF KASHMIR
I SEHIES

I)

BY

1:3. O. COVENTRY, F.C. H.,

WIT II

1l1:~UlII'TIONS

COLtHII<I:D ILLU"T(!ATiONS

or:

ANIl
I'II'T\, SPEUES

1< 1:1 'I<OllllCED PIIOM
IHIIELT COLl)1 III PliOT()( ;I<AI'IIS.

< ,

r'''_
kwf.txt
```

```
[vagrant@c7401 ~]$ egrep -n --color China kwf.txt
877:China, N. Africa.
2445:exported in this form to China, wlicre the product is
2457:root usec!nutside of China and JapfLl1 is insignilicant.
3394:Bhotan; China and Japan.
3509:Great Britain), N. Asia, China, Japan, Java.
3569:N. Asia, China, Japan, Australia, New Zealand.
```

Next, I moved to the step of extracting the page images and creating a contact sheet. I then created an image directory to dump all the images into. We then found out that the images were not all pictures. Some were text and the image we texted did not display because the server I was connected to was unresponsive. However, the image was a text image and not a flower.

```
Syntax Error (3555922): Unknown segment type in JBIG2 stream
Syntax Error (3587219): Unknown segment type in JBIG2 stream
Syntax Error (3587978): Unknown segment type in JBIG2 stream
Syntax Error (3605719): Unknown segment type in JBIG2 stream
Syntax Error (3643031): Unknown segment type in JBIG2 stream
Syntax Error (3665276): Unknown segment type in JBIG2 stream
Syntax Error (3695371): Unknown segment type in JBIG2 stream
Syntax Error (3716012): Unknown segment type in JBIG2 stream
Syntax Error (3716898): Unknown segment type in JBIG2 stream
Syntax Error (3728786): Unknown segment type in JBIG2 stream
Syntax Error (3752674): Unknown segment type in JBIG2 stream
Syntax Error (3763434): Unknown segment type in JBIG2 stream
Syntax Error (3775271): Unknown segment type in JBIG2 stream
Syntax Error (3794277): Unknown segment type in JBIG2 stream
Syntax Error (3810077): Unknown segment type in JBIG2 stream
Syntax Error (3831481): Unknown segment type in JBIG2 stream
Syntax Error (3841438): Unknown segment type in JBIG2 stream
Syntax Error (3856211): Unknown segment type in JBIG2 stream
Syntax Error (3874517): Unknown segment type in JBIG2 stream
Syntax Error (3892742): Unknown segment type in JBIG2 stream
```

```
[vagrant@c7401 ~]$ ls images
KashmirWildflowers-000.pbm  KashmirWildflowers-031.pbm  KashmirWildflowers-062.pbm  KashmirWildflowers-093.pbm  KashmirWildflowers-124.pbm  KashmirWildflowers-155.pbm
KashmirWildflowers-001.pbm  KashmirWildflowers-032.pbm  KashmirWildflowers-063.pbm  KashmirWildflowers-094.pbm  KashmirWildflowers-125.pbm  KashmirWildflowers-156.pbm
KashmirWildflowers-002.pbm  KashmirWildflowers-033.pbm  KashmirWildflowers-064.pbm  KashmirWildflowers-095.pbm  KashmirWildflowers-126.pbm  KashmirWildflowers-157.pbm
KashmirWildflowers-003.pbm  KashmirWildflowers-034.pbm  KashmirWildflowers-065.pbm  KashmirWildflowers-096.pbm  KashmirWildflowers-127.pbm  KashmirWildflowers-158.pbm
KashmirWildflowers-004.pbm  KashmirWildflowers-035.pbm  KashmirWildflowers-066.pbm  KashmirWildflowers-097.pbm  KashmirWildflowers-128.pbm  KashmirWildflowers-159.pbm
KashmirWildflowers-005.pbm  KashmirWildflowers-036.pbm  KashmirWildflowers-067.pbm  KashmirWildflowers-098.pbm  KashmirWildflowers-129.pbm  KashmirWildflowers-160.pbm
KashmirWildflowers-006.pbm  KashmirWildflowers-037.pbm  KashmirWildflowers-068.pbm  KashmirWildflowers-099.pbm  KashmirWildflowers-130.pbm  KashmirWildflowers-161.pbm
KashmirWildflowers-007.pbm  KashmirWildflowers-038.pbm  KashmirWildflowers-069.pbm  KashmirWildflowers-100.pbm  KashmirWildflowers-131.pbm  KashmirWildflowers-162.pbm
KashmirWildflowers-008.pbm  KashmirWildflowers-039.pbm  KashmirWildflowers-070.pbm  KashmirWildflowers-101.pbm  KashmirWildflowers-132.pbm  KashmirWildflowers-163.pbm
KashmirWildflowers-009.pbm  KashmirWildflowers-040.pbm  KashmirWildflowers-071.pbm  KashmirWildflowers-102.pbm  KashmirWildflowers-133.pbm  KashmirWildflowers-164.pbm
KashmirWildflowers-010.pbm  KashmirWildflowers-041.pbm  KashmirWildflowers-072.pbm  KashmirWildflowers-103.pbm  KashmirWildflowers-134.pbm  KashmirWildflowers-165.pbm
KashmirWildflowers-011.pbm  KashmirWildflowers-042.pbm  KashmirWildflowers-073.pbm  KashmirWildflowers-104.pbm  KashmirWildflowers-135.pbm  KashmirWildflowers-166.pbm
KashmirWildflowers-012.pbm  KashmirWildflowers-043.pbm  KashmirWildflowers-074.pbm  KashmirWildflowers-105.pbm  KashmirWildflowers-136.pbm  KashmirWildflowers-167.pbm
KashmirWildflowers-013.pbm  KashmirWildflowers-044.pbm  KashmirWildflowers-075.pbm  KashmirWildflowers-106.pbm  KashmirWildflowers-137.pbm  KashmirWildflowers-168.pbm
KashmirWildflowers-014.pbm  KashmirWildflowers-045.pbm  KashmirWildflowers-076.pbm  KashmirWildflowers-107.pbm  KashmirWildflowers-138.pbm  KashmirWildflowers-169.pbm
KashmirWildflowers-015.pbm  KashmirWildflowers-046.pbm  KashmirWildflowers-077.pbm  KashmirWildflowers-108.pbm  KashmirWildflowers-139.pbm  KashmirWildflowers-170.pbm
KashmirWildflowers-016.pbm  KashmirWildflowers-047.pbm  KashmirWildflowers-078.pbm  KashmirWildflowers-109.pbm  KashmirWildflowers-140.pbm  KashmirWildflowers-171.pbm
KashmirWildflowers-017.pbm  KashmirWildflowers-048.pbm  KashmirWildflowers-079.pbm  KashmirWildflowers-110.pbm  KashmirWildflowers-141.pbm  KashmirWildflowers-172.pbm
KashmirWildflowers-018.pbm  KashmirWildflowers-049.pbm  KashmirWildflowers-080.pbm  KashmirWildflowers-111.pbm  KashmirWildflowers-142.pbm  KashmirWildflowers-173.pbm
```

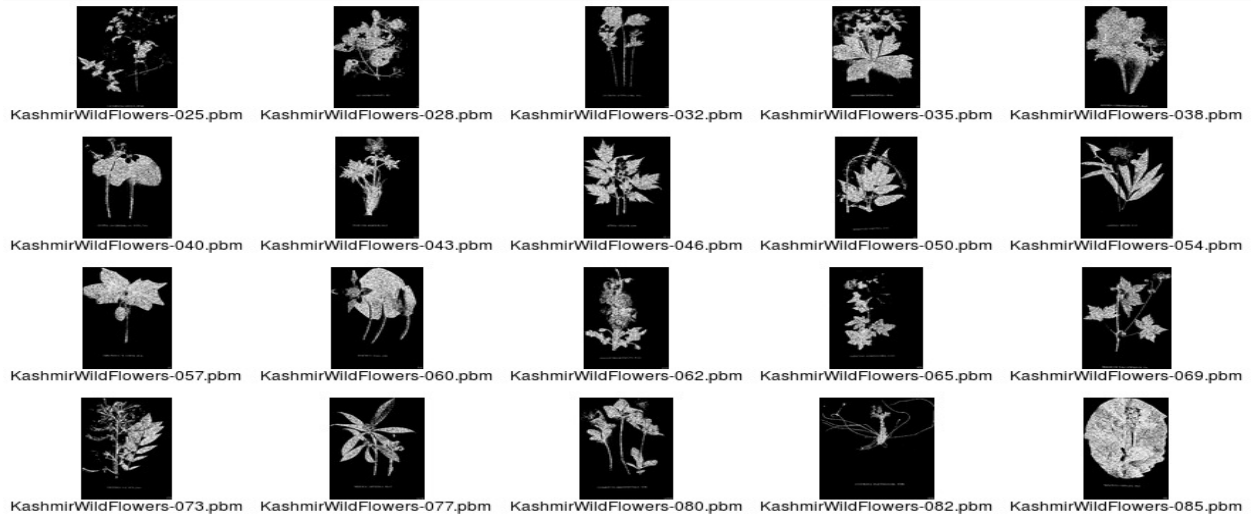
I then used the image magic command to comprise a page of the actual images, a contact list of the thumbnail images.

```
[vagrant@c7401 ~]$ images/KashmirWildflowers-000.pbm[100x100]=>images/KashmirWildflowers-000.pbm PBM 864x928=>93x100 93x100+0+0 1-bit Bilevel DirectClass 0.060u 0:00.06
0
images/KashmirWildflowers-001.pbm[100x100]=>images/KashmirWildflowers-001.pbm PBM 1008x1664=>61x100 61x100+0+0 1-bit Bilevel DirectClass 0.120u 0:00.139
images/KashmirWildflowers-002.pbm[100x100]=>images/KashmirWildflowers-002.pbm PBM 864x1680=>51x100 51x100+0+0 1-bit Bilevel DirectClass 0.100u 0:00.100
images/KashmirWildflowers-003.pbm[100x100]=>images/KashmirWildflowers-003.pbm PBM 16x512=>3x100 3x100+0+0 1-bit Bilevel DirectClass 0.010u 0:00.000
images/KashmirWildflowers-004.pbm[100x100]=>images/KashmirWildflowers-004.pbm PBM 976x1488=>66x100 66x100+0+0 1-bit Bilevel DirectClass 0.110u 0:00.110
images/KashmirWildflowers-005.pbm[100x100]=>images/KashmirWildflowers-005.pbm PBM 1024x1648=>62x100 62x100+0+0 1-bit Bilevel DirectClass 0.110u 0:00.120
images/KashmirWildflowers-006.pbm[100x100]=>images/KashmirWildflowers-006.pbm PBM 960x1632=>59x100 59x100+0+0 1-bit Bilevel DirectClass 0.110u 0:00.110
images/KashmirWildflowers-007.pbm[100x100]=>images/KashmirWildflowers-007.pbm PBM 1008x1648=>61x100 61x100+0+0 1-bit Bilevel DirectClass 0.120u 0:00.120
images/KashmirWildflowers-008.pbm[100x100]=>images/KashmirWildflowers-008.pbm PBM 944x1456=>65x100 65x100+0+0 1-bit Bilevel DirectClass 0.100u 0:00.099
images/KashmirWildflowers-009.pbm[100x100]=>images/KashmirWildflowers-009.pbm PBM 16x16=>100x100 100x100+0+0 1-bit Bilevel DirectClass 0.000u 0:00.000
images/KashmirWildflowers-010.pbm[100x100]=>images/KashmirWildflowers-010.pbm PBM 944x1424=>66x100 66x100+0+0 1-bit Bilevel DirectClass 0.100u 0:00.089
images/KashmirWildflowers-011.pbm[100x100]=>images/KashmirWildflowers-011.pbm PBM 1008x1664=>61x100 61x100+0+0 1-bit Bilevel DirectClass 0.120u 0:00.120
images/KashmirWildflowers-012.pbm[100x100]=>images/KashmirWildflowers-012.pbm PBM 960x1648=>58x100 58x100+0+0 1-bit Bilevel DirectClass 0.110u 0:00.109
images/KashmirWildflowers-013.pbm[100x100]=>images/KashmirWildflowers-013.pbm PBM 976x1664=>59x100 59x100+0+0 1-bit Bilevel DirectClass 0.120u 0:00.110
images/KashmirWildflowers-014.pbm[100x100]=>images/KashmirWildflowers-014.pbm PBM 1008x1680=>60x100 60x100+0+0 1-bit Bilevel DirectClass 0.120u 0:00.120
images/KashmirWildflowers-015.pbm[100x100]=>images/KashmirWildflowers-015.pbm PBM 976x1696=>58x100 58x100+0+0 1-bit Bilevel DirectClass 0.110u 0:00.120
images/KashmirWildflowers-016.pbm[100x100]=>images/KashmirWildflowers-016.pbm PBM 960x1664=>58x100 58x100+0+0 1-bit Bilevel DirectClass 0.110u 0:00.120
images/KashmirWildflowers-017.pbm[100x100]=>images/KashmirWildflowers-017.pbm PBM 992x1664=>60x100 60x100+0+0 1-bit Bilevel DirectClass 0.120u 0:00.120
images/KashmirWildflowers-018.pbm[100x100]=>images/KashmirWildflowers-018.pbm PBM 960x1680=>57x100 57x100+0+0 1-bit Bilevel DirectClass 0.130u 0:00.129
```

I then created a single contact image sheet with the selected images from the code in the example. It created a pdf to export.

```
[vagrant@c7401 ~]$ filenums=(025 028 032 035 038 040 043 046 050 054 057 060 062 065 069 073 077 080 082 085 088 091 094 096 098)
[vagrant@c7401 ~]$ for num in ${filenums[@]} ; do echo images/KashmirWildflowers-${num}.pbm ; done
images/KashmirWildflowers-025.pbm
images/KashmirWildflowers-028.pbm
images/KashmirWildflowers-032.pbm
images/KashmirWildflowers-035.pbm
images/KashmirWildflowers-038.pbm
images/KashmirWildflowers-040.pbm
images/KashmirWildflowers-043.pbm
images/KashmirWildflowers-046.pbm
images/KashmirWildflowers-050.pbm
images/KashmirWildflowers-054.pbm
images/KashmirWildflowers-057.pbm
images/KashmirWildflowers-060.pbm
images/KashmirWildflowers-062.pbm
images/KashmirWildflowers-065.pbm
images/KashmirWildflowers-069.pbm
images/KashmirWildflowers-073.pbm
images/KashmirWildflowers-077.pbm
images/KashmirWildflowers-080.pbm
images/KashmirWildflowers-082.pbm
images/KashmirWildflowers-085.pbm
images/KashmirWildflowers-088.pbm
images/KashmirWildflowers-091.pbm
images/KashmirWildflowers-094.pbm
images/KashmirWildflowers-096.pbm
images/KashmirWildflowers-098.pbm
```

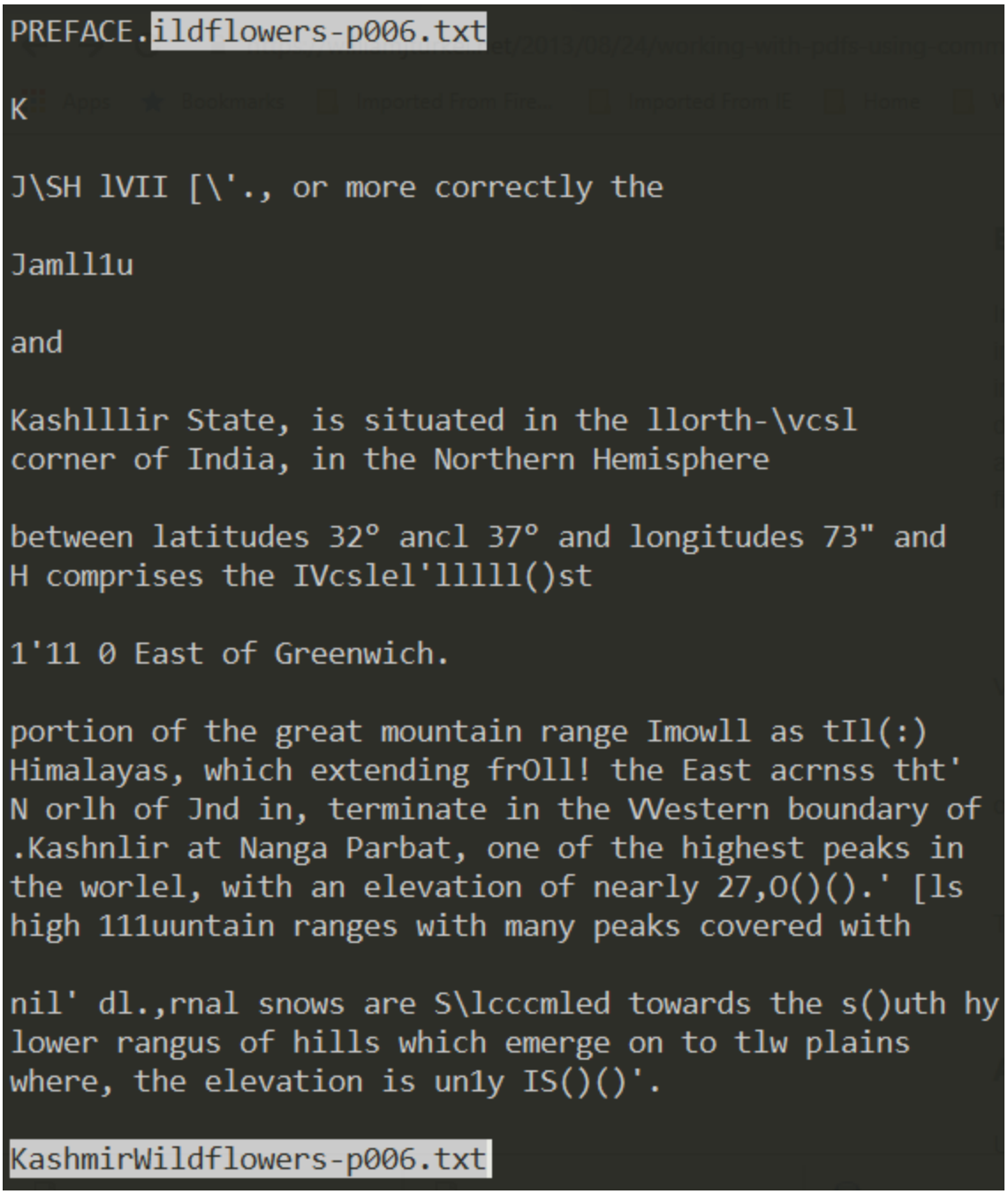
The final pdf file with the select images came out as a grid. I took a snapshot of a piece of the page.



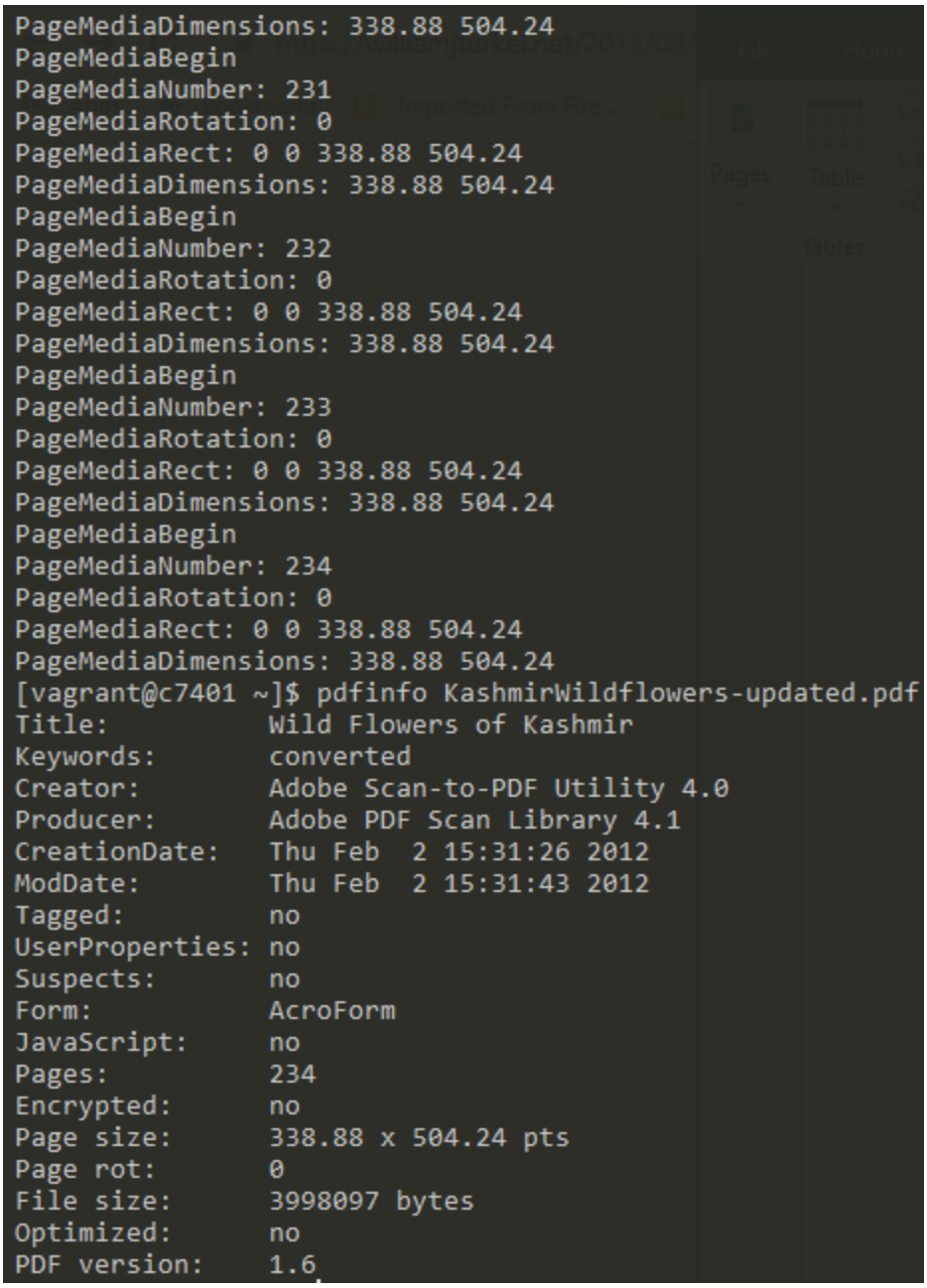
I then exported the file to a pdf after manipulating it with the pdftk command.

```
[vagrant@c7401 ~]$ pdfinfo KashmirWildflowers.pdf
Keywords:       converted
Creator:        Adobe Scan-to-PDF Utility 4.0
Producer:       Adobe PDF Scan Library 4.1
CreationDate:   Thu Feb  2 15:31:26 2012
ModDate:        Thu Feb  2 15:31:43 2012
Tagged:         no
UserProperties: no
Suspects:       no
Form:           AcroForm
JavaScript:     no
Pages:          234
Encrypted:      no
Page size:      338.88 x 504.24 pts
Page rot:       0
File size:      3945597 bytes
Optimized:      yes
PDF version:    1.6
```

I then made sure I burst the file into smaller ones since it is one that is very long. So, we used the pdftk command to make the large doc into a separated single pages that way we can search them better. Then a metadata file was created and a new txt file. I checked my work with looking at page 3 once the burst was complete. I then turned that specific image into a text document like we did above. I then made sure by looking at it through the less command.



As we can see it was a success converting the PDF to a text file. We then wanted to explore how to add information to pdfs once we complete new txt and pdf files. We created a new file with the information we wanted to add to it. Once that was done we updated the file and made a new pdf.



The above is the final pdf that is created with the modified data in it.

Analysis of Results

We see that we can make and read pdfs with the above commands and packages in Linux. This is a great way to modify them and still have them come out the way we would want instead of having to create a completely new pdf. We see that we can add to the pdf, take away from it, as well as move things

around in it. We can also create new different pdfs with new information in them while including all the old information. But most of all we found a way to be able to manipulate images from pdfs, jpegs, pbms, and more by turning it to text and then manipulating it before turning it back to an image. This is a truly amazing ability.

Conclusion

Overall, I wanted to show the use of Linux and how it can be very beneficial in the manipulation of pdfs and images. I think that this will come in handy for future use, because There will be times in which we need to change or manipulate and image without being able to physically change it or create a new one. This program could allow us to rapidly change the pdf and image by turning it into text and then allowing changes before creating a new one or updating the old one.

Resources:

1. Assignment and commands. https://williamjturkel.net/2013/08/24/working-with-pdfs-using-command-line-tools-in-linux/
2. Linux help. https://www.linuxglobal.com/pdftk-works-on-centos-7/
3. Centos 7 packages. https://centos.pkgs.org/7/epel-x86_64/xpdf-3.04-9.el7.x86_64.rpm.html
4. Centos 7 downloads. https://download-ib01.fedoraproject.org/pub/epel/7/x86_64/Packages/e/
5. Book archive.gov. https://ia801602.us.archive.org/21/items/WildFlowersOfKashmir/KashmirWildflowers.pdf