FinalProject Faillog

1: Transciption

With the help of classmates, and some extra work using BBEDIT on my computer I was able to transcribe the files. It took about 5hrs to get it to where I wanted it to be. I edited out information I felt might not be necessary for the conversation I wanted to start about the diary texts. I started this project by feeling incredibly confused as to the direction I wanted to take, and I had to almost force myself to sit and do the work. You can say I was afraid of the unknown.

Once I was able to transcribe the files in a way that I thought was easy enough for me to work with, I started thinking of all the possible tools I could use to make sense of it. From the start, I had this idea of making sense of the word that popped out to me the most which was "strength" but once I got into the transcriptions and with the help of the work my classmates Lauren had done, I was more interested in finding connections between dates and the events. I wanted to create a CSV file so I could use it for Text Analysis and Topic Modeling but for so many reasons, things weren't working out for me. So I decided to go through each line in BBEdit and separate the lines with commas and instead of grouping texts by sections, I grouped by place. Labeling each line that had an event occurring in e.g., London, London. Ending up with three columns

- Place

- Date

- Event

After a couple of hours, I was able to get a CSV file I could work with.

2: Open Refine

Once I was able to create a decent CSV file, I uploaded it into OpenRefine to clean it out a bit more. Once I opened the file to clean it up, everything started making a bit more sense and following the tutorial from Exercise 3 made things run faster. Saved the new CSV file and uploaded it into my GitHub. This CSV helped me with the next program. I realized once I got here, that I could've done the Place, Data and Event headings in OpenRefine rather than by hand. So in my final CSV file, you'll see two headings. For the following exercise though, I went in and erased line 1 which I had originally created by hand.

3: Gehpi & Voyant Tools

Prior to doing the final project, I was so sure Voyant Tools was out of the question, seeing as it never worked well for me before. But I decided to try it out and see if I can make meaning out of it. So I went back to the tutorials and followed step by step, keeping my fingers crossed the program would work in my favor. And voila! it did. Honestly was a bit of a miracle as I was getting really tired at this point. Uploaded my CSV file into Voyant tools to visualize connections. I found out that by cleaning the data by hand, I had altered my results. The most frequent word was now **Eastbourne** and **Sussex** occuring 117 and 116 times respectively, because it was the most frequent location in my cleaned text file. If i hadn't grouped my data the way I did, the most frequent would've been hospital; occuring 64 times. The graph analysis also supported the findings. I set **stopwords** to English and I selected apply globally. I decided to look at the wordtree connections and it created branches that gave a bit more explanation to the most frequent words. For instance, hospitals were located in Eastbourne, saints and Canadian military went to the hospitals in

Eastbourne and there were labs and theatres in these hospitals. It also showed that Eastbourne was equipped and re-named. So you either proceed to the Canadian Military Hospital, or it's renamed Canadian Military Hospital was present in the diaries.

I imported my spreadsheet again into Gephi for a different type of visual. Making sure I selected Comma and the As table is listed as Edges table. Gephi is a bit of a mess to me. I had to manipulate the tools; Overview pane, Data laboratory and Preview pane, on there a bit - reducing the text size, rescaling the weight, proportional size a bit more than was done in the tutorial in Exercise 4. The output was slightly clearer than when I first did it and one thing stood out clearly; **the strength of the unit unchanged** was quite interesting to see especially thinking of the period when these files were gathered. The top also talked about troop hospital on board taken and a rank unit transfer which i'm guessing was talking about an ranked officer being admitted in a hospital for them. One thing i thought might change things is uploading the CSV file with the headings: place, date and event. Could have been a bit clearer if I didn't. Also figured that if I was an expert on this period in history, the analysis would make more clear.

4: Topic Modeling

I really struggled with topic modeling in R. Was having the worst luck here, with RStudio on my laptop and in DH Box. From my last fail log I had mentioned I struggled with unpacking rJAva into RStudio, the same problem persisted and I was unable to use the RStudio on my laptop.

Error:

> library(rJava)

Error: package or namespace load failed for 'rJava':

.onLoad failed in loadNamespace() for 'rJava', details:

call: dyn.load(file, DLLpath = DLLpath, ...)

error: unable to load shared object

'/Library/Frameworks/R.framework/Versions/3.5/Resources/library/rJava/lib

dlopen(/Library/Frameworks/R.framework/Versions/3.5/Resources/library/rJ

o, 6): Library not loaded: /Library/Java/JavaVirtualMachines/jdk-

9.jdk/Contents/Home/lib/server/libjvm.dylib

Referenced from:

/Library/Frameworks/R.framework/Versions/3.5/Resources/library/rJava/l

ibs/rJava.so

Reason: image not found*

I tried the RStudio in the DH Box which worked really well for me the in Module 4 but that was a disaster too. I figured it must have something to do with my CSV file. I also couldn't find the Environment and History section of RStuido in the DHBox; where I was able to see all the packages unpacked and the files uploaded. The issue persisted when I tried to tell Mallet how many topics to search for.

*> num.topics <- 20

> topic.model <- MalletLDA(num.topics)
Jun 20, 2018 11:19:09 PM cc.mallet.topics.ParallelTopicModel <init>
INFO: Coded LDA: 20 topics, 5 topic bits, 11111 topic mask
topic.model$loadDocuments(mallet.instances)
Error in lapply(list(...), ._java_valid_object) :
object 'mallet.instances' not found

> num.topics <- 10
>
> topic.model <- MalletLDA(num.topics)
>
> Jun 20, 2018 11:20:57 PM cc.mallet.topics.ParallelTopicModel <init>
>
> INFO: Coded LDA: 10 topics, 4 topic bits, 1111 topic mask
>
> topic.model$loadDocuments(mallet.instances)
>
> Error in lapply(list(...), ._java_valid_object) :
>
> object 'mallet.instances' not found
>
> I got to the end of what i call my final project only to realize, i needed to unpack rJava in RStudio which is why i kept getting the error above as well as the one below:

*> topic.model <- MalletLDA(num.topics)

Jun 21, 2018 12:36:57 AM cc.mallet.topics.ParallelTopicModel <init>

INFO: Coded LDA: 20 topics, 5 topic bits, 11111 topic mask

> topic.model$loadDocuments(mallet.instances)
>
> Error in lapply(list(...), ._java_valid_object) :
>
> object 'mallet.instances' not found
>
> vocabulary <- topic.model$getVocabulary()
>
> Error in .jcall("RJavaTools", "Ljava/lang/Object;", "invokeMethod", cl, :
>
> java.lang.NullPointerException
>
> word.freqs <- mallet.word.freqs(topic.model)
>
> Error in .jcall("RJavaTools", "Ljava/lang/Object;", "invokeMethod", cl, :
>
> java.lang.IllegalStateException: You must load instances before you can count features

Once I got really fed up of working in RStudio, I decided to use the TopicModeling Tool instead which was actually quite straight forward. I set

the number of topics I wanted to model to: 10 at first, then i decided to try **15** instead. I clicked on train topics which ran the alogrithim and saved two folders on my laptop. Words like hospital, war, nursing, strength were expected.

5: RAW

Since my file was pretty clean from hand transcription, I decided to upload it straight into Raw to create charts. I picked Place and Event and created an alluvial diagram. Was quite splendid. I was impressed thinking of where I was coming from. You could see connections like **Eastbourne being renamed Canadian Military Hospital. Dispatch and Ambulance Service twice a day between Eastbourne and Seaford (nine miles). Authority received from A.D.M.S Brighton to use ward in Seaside Convalescent Hospital SEAFORD as auxiliary. I also looked at a denogram. Again I was quite pleased with the results. It has to be the cleanest visual I have created so far in this course. Showed clean connection between the date, place and events occurring.

**All files for this final project can be found in my github.