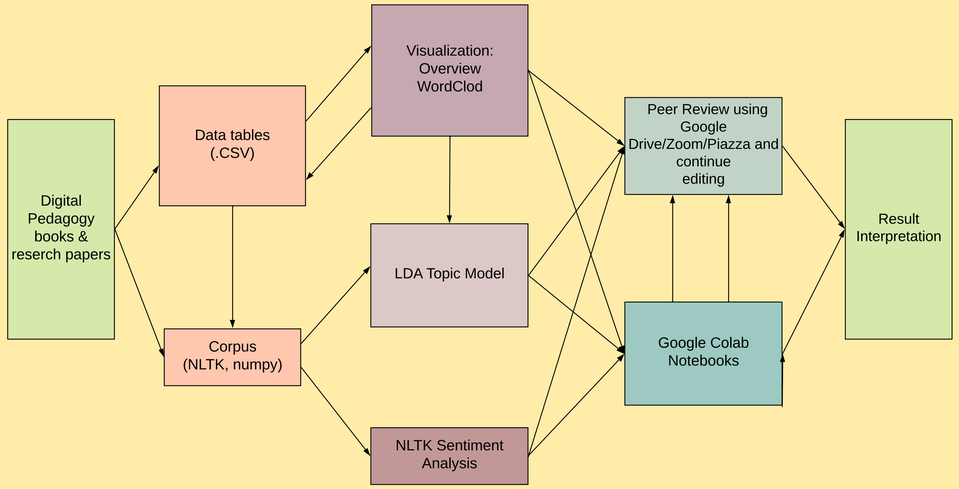
Digital Pedagogy Exploration

DH 100:
Theory and Method in Digita Humanity

Instructor: Dr. Anderson

Student: Sufang Tan



Descriptions:

Project:
Over the past decade, online education and classroom-based education have begun to converge in the form of digital pedagogy. Recently, due to the impact of Covid-19, distance education has embedded in every educator's daily life. It is interesting to understand the impact of the disruptive force of technology to 21st-century education. This analysis is aimed to have a more comprehensive understanding of how digital pedagogy is addressing issues occurring in traditional education and potential problems it brings to.

Dataset :
The dataset consists of the texts related to digital pedagogy from more than 50 e-books, on average, having 200 pages, and some researcher papers written by professionals in this field. Most of them are published in the last decade.

Questions for Exploratory Data Analysis:

1)What is scholars' attitude towards digital education and how it changes through the development of technology and the implementation of that development?
Answer: Using NLTK to extract the adjectives or other keywords in python and then by using TextBlob in Python to classify the texts into two categories, more positive or more negative. Maybe create a timeline of the texts to evaluate changes.

2)What are some effective pedagogical methods to better engaging in digital word? Answer: Create a subset of data content texts with high frequency in possitive adjectives. Using topic modeling tool to generate different topics and analyze each categogy to evaluate the inner connection and the similarity,eg. They implemented the same technology or they help to practice one methodology of pedagogy.

3)What are the pedagogy critiques involved in technology innovation and how to solve potential problems?
Answer: Create a subset of data content texts with high frequency in negative adjectives. Using Word Cloud to generate the most frequent words, and analyzing the context of those negative adjectives to find the potential reasons for the criticism and solutions.

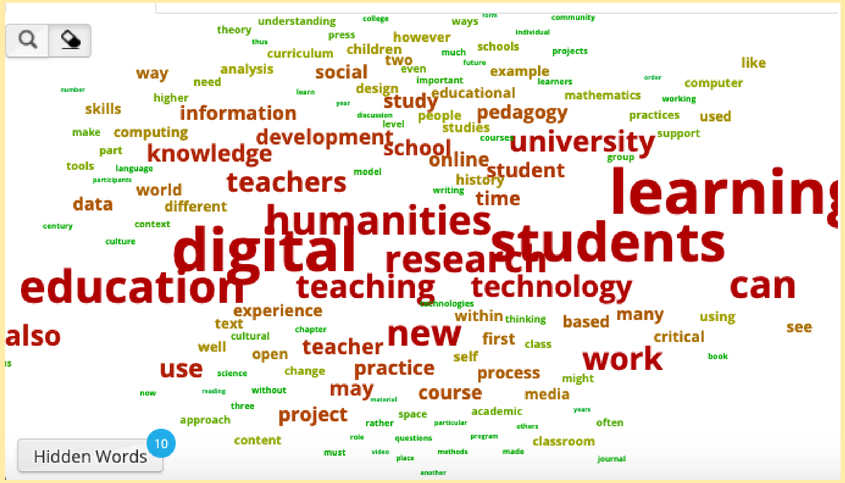
Descriptions of Tools & Methodst :

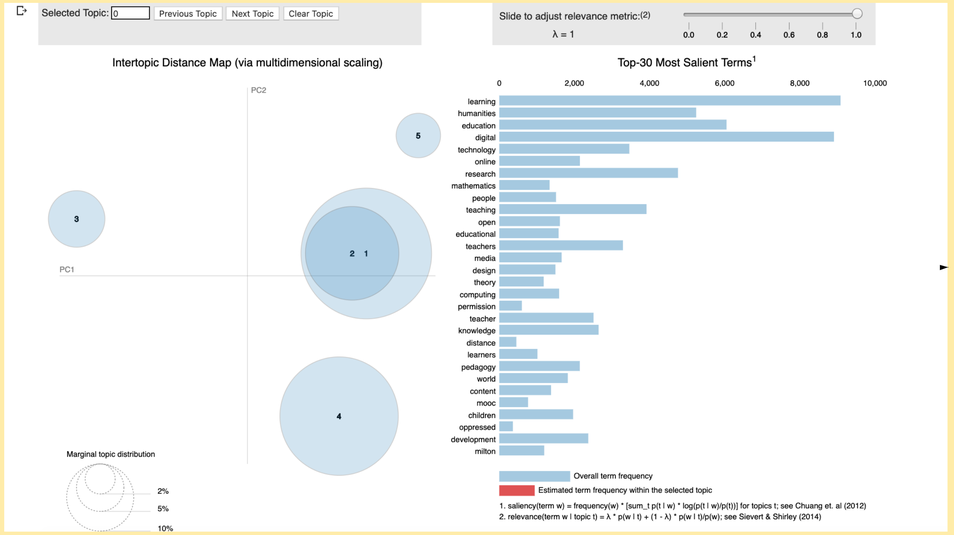
1.Overview Word Cloud: It can generate a word cloud showing a group of the most frequent words in 43 documents. It can help better understand the mean topic in all documents.

2. Text preprocessing (e.g. tokenize, removal stopwords and punctuations, lower capitalize) of scraped text using NLTK (natural languages tool kit).

3. Topic modeling (LDA): the LDA results are assigned as weights for directed edges, with each source as the highest-scoring documents within a topic, to each target drawn from the subsequent highest scoring documents within the same topic. The number of topics can be adjusted within the notebook, as can the number of edges drawn for each topic. Through the experiments, 5 topics could be a reasonable size for my dataset. Because of the limitation of the dataset size, topic 1,2,3 are highly overlapping

4. VADER Sentiment analysis: lexicon sentiment analysis tool that is specifically attuned to sentiments expressed in social media and documents. It adds predetermined scores from NLTK's sentiment analysis package and assigns a positive, negative, neutral and compound rate to each document, however, the real emotion behind the document can not be full express on the rate because its analysis is based on each word instead of the whole sentence structural, even it can deal with the context of the word, some of the articles may use rhetorical or ironic methods, which hardly detected by this tool.

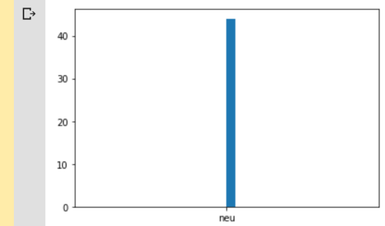





```
import matplotlib
import matplotlib.pyplot as plt

atitudes= ['neu', 'neu', 'neu', 'neu', 'neu', 'neu', 'neu', 'neu',

plt.hist(atitudes,bins = 30);
```



Interpret Result:

The OverviewDocs analysis was able to give me an overview of what my dataset would indicate pedagogy education into Topic Modeling and LDA methodology. OverviewDocs provided a rather surface-level answer to my EDA question regarding the most common aspects of education impacted by the digital world. The entities feature of OverviewDocs proved to be the most useful because it counted the frequency of words, and it can indicate some specific technology tools related to education like computer programs, media.

```
[ ] SOME_FIXED_SEED = 42

    np.random.seed(SOME_FIXED_SEED)

    lda = LDA(n_components=number_topics, n_jobs=-1)
    lda.fit(count_data)

    print("Topics found via LDA:")
    count_topics(lda, count_vectorizer, number_words)
```

```
Topics found via LDA:

Topic #0:
learning education students digital teachers research teaching new technology university

Topic #1:
learning students mathematics education teachers teaching use children work teacher

Topic #2:
education permission people online oppressed world owner copyright pedagogy technology

Topic #3:
digital humanities students research new computing project http text work

Topic #4:
learning education technology distance teaching research open theory educational online
```

LDA :

The LDA Topic Model analyzes the correlation with my dataset documents and to evaluate the potential topic that occurs in the total dataset. A closer view of these data models can be found here. There are three key aspects of the LDA visualization to note: saliency, relevance, and bubble size. Saliency measures how much a particular term can tell you about the topic. Relevance is the " weighted average" of the probability of a word in a particular topic normalized by the probability of the topic. The size of each bubble indicates how important a topic is concerning the dataset based on the frequency of the words within each topic. In the notebook, this model is interaction and users can hover over each topic bubble to see the most salient terms, along with the top 30 most salient terms overall. Furthermore, bubble clustering means similar topics or themes within the books and papers.

The LDA Topic generates 5 meaningful topics from my dataset which covers five different educational areas that improved by technology development, and they all worth having a further investigation. The first one is related to digital pedagogy in university research. The second indicates that the learning and teaching in mathematical education have been influenced by technology, which probably also refers to changing on mathematically related subjects(STEM). In the graph the first two topics are overlapped and the topic 2 is inside the topic 1 that is because of the fact that topic 1 contains all the words in topic 2, and those words are closer clustering with each other inside the topic 1, so it is reasonable to generate a new topic to narrow the scope. The third topic is about digital humanity education by using computational tools to solve humanity's subject problem and deeply analyze text-based subjects (history/media/ literature). Last but not the least, topic 5 focus on online education and other new education formats that solve the distant education that helps education resource more reachable for all learners. This topic's bubble has the smallest size but it does not imply topic 5 is the least popular topic or having a trivial effect, indeed, it has been ever closer related to our life today, the period of social distance restriction. Since my dataset size is limited to 50 documents and the contents are not subjectively, this may result in a relatively small proportion of topic 5.

VADER Sentiment analysis:
Surprisingly, all my corpus tested as neutrality. However, this result is also understandable since VADER extract emotional or sentiment polarity from text to matching with its lexicon, it will perform well on text with explicit emotional words like social media text, while my dataset is all research papers or books which usually using neutral language and less emotional. Moreover, it is also reasonable that most of the authors hold a neutral standpoint about the influence of technology on education.

Links& References:

1)
DH100 Class Resources:
https://drive.google.com/drive/folders/1fWQvXlU-4HL3vGDvfHG0-E7kg8X3n-97

2)
Overview(wordCloud)
https://www.overviewdocs.com/documentsets/22082/view-33045478414701

3)
Oskicat Berkeley Library
http://oskicat.berkeley.edu/search

4)
Google Schooler
https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=digital+pedagogy&btnG=

5)
Make_csv_tutorial, 07/20/2020
https://colab.research.google.com/drive/1p_-zC9aY112ioXFGArcbTMt-vdKaj7Gw

6)
PaulShao, WordClouds_TopicModels Notebook , 07/02/2020
https://colab.research.google.com/drive/1mGf_RZaxOY-Sutbwzm0kJMknql5cGYin