Final Project

Ziyan Zhuang

CL4HSS 2024

Themes discussed under *r/Buddhism* on *Reddit*

1. Introduction

The aim of this project is to explore online Buddhist community on Reddit - an American forum social network and the 9th most-visited website in the world (Wikipedia Contributors, 2019) - using means of distant reading. Related work like *Distant Reading of Religious Online Communities: A Case Study for Three Religious Forums on Reddit* gives an example of implementing distant reading on three online communities with subreddit *Christianity, Islam* and *Occult*, with computational methods including bigram frequencies, collocations, topic modeling and sentiment analysis (Schmidt et al., 2020). Analysis of Buddhist communities on *Reddit* is missing so the research question this time is to find out what themes are discussed under the subreddit *r/Buddhism*.

2. Steps of process

a) Grasp and save the data in the CSV format in a database.

b) Break down sentence into tokens using tokenizing tools, and clean up empty lines and replicates.

c) Calculate single word frequency.

d) Calculate the frequency of word sets to reveal the context of the words.

e) Detect possible topics.

3. Data, tools and methods in each step

Throughout the project, I used python as programming language and VS Code as code editor.

3.1 In ‘step a’

The raw data I got in previous ‘step a’ are posts and comments under the subreddit *r/Buddhism* on *Reddit.* I imported the Python Reddit API Wrapper (PRAW) library, aided with Pandas Python library and Python os Module. I first tried to get the data sorted by hot

(sortbyhot.py), which is the default sorting method, but I found out that the results could vary within a day due to the changing feature of the algorithm. Therefore, I switched to sort by top and this year (sortbytopyear.py) and added Python time Module, leading to a more stable result. I saved the results to 'buddhism_posts_incremental.csv' (200 posts) and 'buddhism_comments_incremental.csv' (13395 comments). Also, I had the ids of posts stored in 'processed_ids.txt' so that I can continue to obtain data where is left off if I need more.

One problem I had is that some posts do not have textual content but have a picture or meme instead. I decided to also capture those graphic data too. I created a field called 'media_url' and saved the pictures as URLs in it.

3.2 In 'step b'

In 'step b', I first tokenized posts data in 'buddhism_posts_incremental.csv' and comments data in 'buddhism_comments_incremental.csv', and then tried to tokenize the texts in the images in 'media_url' field of 'buddhism_posts_incremental.csv'.

Posts data have two fields that have the text I wish to tokenize - 'title' and 'selftext', while comments data only have one field - 'body'. The codes in 'tokenize_posts_topyear.py' and 'tokenize_comments_topyear.py' do the job respectively. I uesed Pandas Python library and NLTK (Natural Language Toolkit). Tokenizing texts in images is more complicated. I used Python-tesseract - an optical character recognition (OCR) tool for python - to extract the text (if any) in the images. And then follow the regular path for tokenizing text. At first tesseract did not work properly and I tried several times to figure out the cause of problem - for some reason, I should install it in my Windows system instead of install in Ubuntu.

The results of 'step b' are stored in 'tokenized_buddhism_posts', 'tokenized_buddhism_comments' and 'tokenized_buddhism_media_texts'.

3.3 In 'step c''

The major tools used here are WordNetLemmatizer, NLTK and regex. The idea is to calculate the frequency of meaningful words in the previous results. First, I merged the three files in 'step b'. Then I downloaded the English stop words from NLTK, such as pronouns and prepositions, to make sure those words would not be calculated. I also used regular expression '^[a-zA-Z]+$' to make sure the words start with one or more uppercase or lowercase letters and also end with an uppercase or lowercase letter, to avoid forms like ''s' and '--'. The result

is saved as ‘word_frequencies.csv’

3.4 In ‘step d’

The essential tool used is N-grams imported from NLTK. The previous step roughly gives a list of single words’ frequency from high to low. However, it does not give any context to the those words, nor high frequency phrases. The use of N-grams can help achieve this. Stop words and regex used in ‘step c’ are also used in this step.

The results of this step are stored as ‘cleaned_unigram_frequencies.csv’, ‘cleaned_bigram_frequencies.csv’, ‘cleaned_trigram_frequencies.csv’.

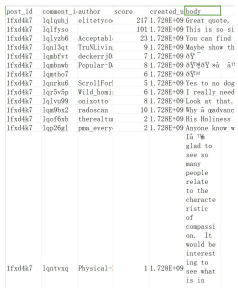
3.5 In ‘step e’

The key in this step is to apply Latent Dirichlet Allocation (LDA) to uncover topical structures. I first installed Gensim library in VS Code terminal. Apart from stops works in NLTK and regex, I also included additional_stop_words, which are "removed", "http", "https", "deleted", "comment", "post", "rule", "nice","cool", "lol". I adjusted the list a few times and I will explain the choice of additional stop words in the ‘Results’ section. The model and dictionary are saved as ‘lda_model_buddhism’ and ‘lda_dictionary_buddhism’.

4. Results and analysis

Apart from ‘lda_model_buddhism’ and ‘lda_dictionary_buddhism’, other outputs are CSV files with encoding utf-8. For presentation, I will open them in WPS excel. Some of the characters cannot be normally presented but it does not hinder later process.

id	title	author	score	url	media_url	num_comme	created_u	selftext
1fxd4k7	Advanced	algreen58	1924	https://i.i	https://i.	101	1.728E+09	
18yep21	Interesti	Professio	1667	https://i.i	https://i.	124	1.704E+09	I know meme
1gwslo7	No one is	Outrageou	1523	https://www.reddit.		104	1.732E+09	These image
1g46pva	Buddhism	Outdoorsy	1491	https://i.i	https://i.	63	1.729E+09	
1dds1ki	I tempora	jacklope	1473	https://www.reddit.		205	1.718E+09	Was given the name Muditananda, one who attains the highest JOY by The Ven Dr Saccanananda Mahathera at Dhamma



post_id	comment_i	author	score	created_u	body
1fxd4k7	lq1quhj	elitetyco	217	1.728E+09	Great quote,
1fxd4k7	lqlfyso		101	1.728E+09	This is so si
1fxd4k7	lqlyzb6	Acceptabl	23	1.728E+09	You can find
1fxd4k7	lqnl3qt	TruNLivin	9	1.728E+09	Maybe show th
1fxd4k7	lqmbfvt	deckerrj0	7	1.728E+09	ðŸ˜
1fxd4k7	lqmbowb	Popular-D	8	1.728E+09	ðŸ™ŒðŸ˜ ðŸ
1fxd4k7	lqmzho7		6	1.728E+09	ðŸ™Œ
1fxd4k7	lqnzkn6	ScrollFor	3	1.728E+09	Yes to no dog
1fxd4k7	lqr5v3p	Wild_homi	6	1.728E+09	I really need
1fxd4k7	lqlvu99	coinotto	8	1.728E+09	Look at that.
1fxd4k7	lqm0to2	radoscan	10	1.728E+09	Why a qndvnc
1fxd4k7	lqof6sb	therealtu	2	1.728E+09	His Holiness
1fxd4k7	lqn26gl	pms_every	2	1.728E+09	Anyone know w
1fxd4k7	lqntvsq	Physical-	1	1.728E+09	Ik I'm glad to see so many people relate to the characteristic of compassion. It would be interesting to see what is in

Figure 1. 200 posts (sort: top, this year) Figure 2. 13395 comments (sort: top, this year)

And then I tokenized the field ‘title’ and ‘selftext’ in posts, the field ‘body’ in comments, and the texts in the images contained in ‘media_url’ in posts. The results for tokenization are as follows:

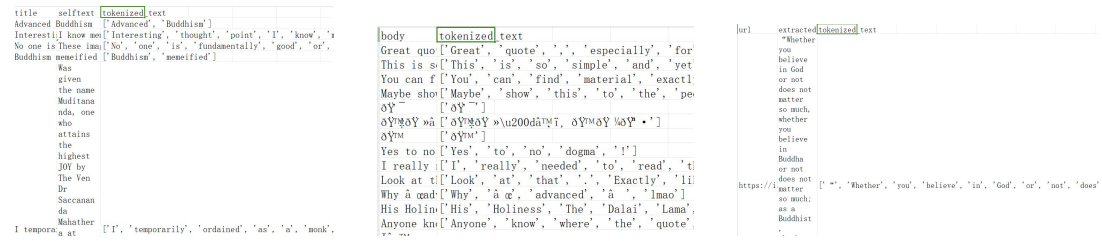


Figure 3-5. tokenized posts, comments and text of images

There are quite a lot of punctuation and abbreviation, as well as numbers, pronouns and other words that do not have significant meanings regarding the topic. Those forms should be excluded for further process.

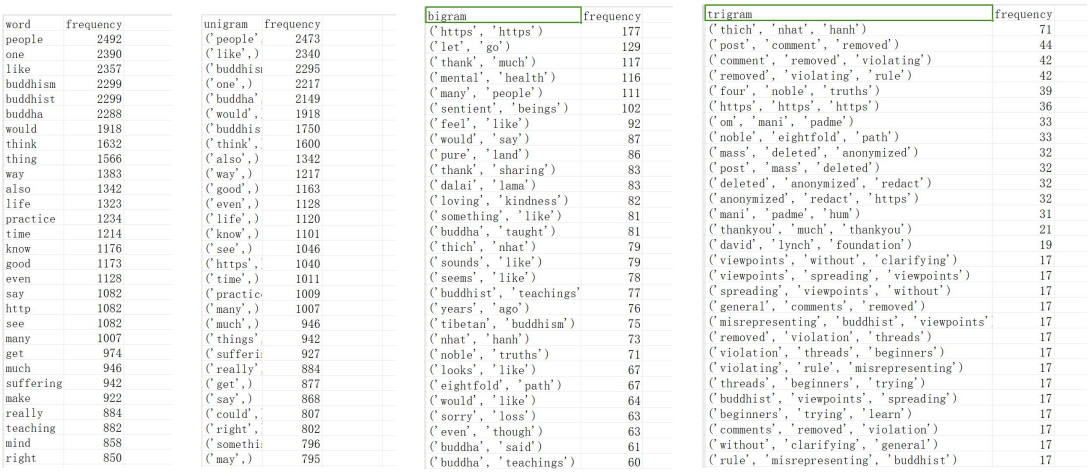


Figure 6: single word frequency (top 30) Figure 7-9: word sets frequency (top 30)

Figure 6 and Figure 7 are both single word frequency, respectively processed by WordNetLemmatizer and N-grams. Though there are words like ‘buddhism’, ‘buddhist’, ‘buddha’, ‘practice’, ‘suffering’, ‘teaching’ and ‘mind’ that point directly at Buddhist belief, practice and teaching, other words are often vague and we cannot imply any possible topics from them. I cannot say what ‘people’, ‘one’, ‘like’, ‘would’, ‘think’ have to offer in terms of topics. I think they rank high because we often use those words to brings things up. They may be close to central topic words, but they themselves are too universal to be topic related. So to discover topics in more details, I need more complicated methods.

In Figure 8 and Figure 9, word pairs give a much clearer picture. In Figure 8, the prominent noise (‘https’, ‘https’) may be caused by people sharing links every now and then. (‘let’, ‘go’), (‘sentient’, ‘being’), (‘noble’, ‘truths’), (‘eightfold’, ‘path’) are connected to basis Buddhist beliefs - after learning the truths of suffering, one should let go to transcend from suffering, one should reach nirvana through eightfold path, and a sentient being is one who perceives or

responds to sensations. It is worth noticing that (‘mental’, ‘health’) ranked fourth, which implies that people in this Reddit Buddhist community put an emphasis on mental health. (‘tibetan’, ‘buddhism’) and (‘pure’, ‘land’) refer to specific traditions of Buddhism. (‘dalai’, ‘lama’), (‘thich’, ‘nhat’), (‘nhat’, ‘hanh’) refer to well known Buddhist figures. Other pairs like (‘buddhist’, ‘teaching’) and (‘buddha’, ‘taught’) are too broad and literal to convey more details, while (‘thank’, ‘much’), (‘thank’, ‘sharing’), (‘many’, ‘people’), (‘feel’, ‘like’), (‘would’, ‘say’) do not have topic specific meanings.

In Figure 9, triple pairs saying a post or comment is removed due to violating the rules are very common noises. Among meaningful pairs, (‘thich’, ‘nhat’, ‘hanh’) is the most used triple pair and is almost twice more frequent than (‘four’, ‘noble’, ‘truths’), which actually ranked the second. This is a surprising finding for me. Thich Nhat Hanh, known as the ‘father of mindfulness’, had a major influence on Western practices of Buddhism (Wikipedia Contributors, 2019b). He founded the Plum Village Tradition, which gave inspiration to engaged Buddhism, and the Plum Village Monastery in France (Wikipedia Contributors, 2019b). (‘om’, ‘mani’, ‘padme’) and (‘mani’, ‘padme’, ‘hum’) are a part of ‘Om mani padme hum’, a well-known Sanskrit mantra. There are also beginners trying to learn Buddhism ((‘threads’, ‘beginners’, ‘trying’) and (‘beginners’, ‘trying’, ‘learn’)). Based on (‘misrepresenting’, ‘buddhist’, ‘viewpoints’) and (‘buddhist’, ‘viewpoints’, ‘spreading’), there could be some debates on Buddhist ideas going on, but possibly they can still be noises. For ‘misrepresenting’ and ‘buddhist’ are also included in (‘rule’, ‘misrepresenting’, ‘buddhist’), which I am not sure whether is a rule violation noise or people discussing Buddhist ideas.

As is shown above, N-grams can provide more details to the topics. However, there are quite a lot of noises, such as (‘https’, ‘https’) and some pairs related to removed comments. And overlapping is also problematic, such as (‘threads’, ‘beginners’, ‘trying’) and (‘beginners’, ‘trying’, ‘learn’). Overlapping is common in Figure 9, because the words in a triple pair are not a pair in natural language, but co-appearing words in a sentence. I want to solve these problems in topic modeling.

```
Top topics with words:
Topic 1: 0.013*"life" + 0.010*"thing" + 0.010*"one" + 0.009*"suffering" + 0.009*"way" + 0.009*"good" + 0.009*"think" + 0.008*"like" + 0.008*"others
" + 0.008*"would"
Topic 2: 0.029*"people" + 0.011*"think" + 0.010*"would" + 0.009*"like" + 0.009*"say" + 0.008*"person" + 0.008*"right" + 0.007*"thing" + 0.007*"poin
t" + 0.007*"u"
Topic 3: 0.023*"buddha" + 0.019*"self" + 0.015*"one" + 0.013*"mind" + 0.013*"path" + 0.010*"suffering" + 0.009*"view" + 0.008*"being" + 0.008*"body
" + 0.007*"teaching"
Topic 4: 0.015*"zen" + 0.010*"west" + 0.010*"parent" + 0.009*"name" + 0.008*"om" + 0.008*"lotus" + 0.008*"hinduism" + 0.007*"seeking" + 0.007*"japa
nese" + 0.006*"nepal"
Topic 5: 0.041*"buddhist" + 0.038*"buddhism" + 0.016*"buddha" + 0.014*"teaching" + 0.011*"many" + 0.010*"religion" + 0.009*"practice" + 0.008*"dhar
ma" + 0.008*"tradition" + 0.008*"monk"
Topic 6: 0.028*"animal" + 0.020*"meat" + 0.012*"killed" + 0.010*"eat" + 0.010*"yes" + 0.010*"nice" + 0.010*"namo" + 0.007*"cool" + 0.007*"kill" + 0
.007*"food"
Topic 7: 0.047*"buddha" + 0.027*"statue" + 0.022*"look" + 0.013*"head" + 0.009*"guru" + 0.009*"thich" + 0.008*"temple" + 0.007*"nhat" + 0.007*"trau
ma" + 0.007*"post"
Topic 8: 0.014*"like" + 0.011*"buddhism" + 0.010*"get" + 0.009*"one" + 0.009*"really" + 0.009*"people" + 0.009*"year" + 0.009*"time" + 0.008*"pract
ice" + 0.007*"know"
Topic 9: 0.041*"book" + 0.028*"read" + 0.011*"know" + 0.010*"one" + 0.010*"would" + 0.008*"good" + 0.008*"also" + 0.008*"zen" + 0.008*"translation"
 + 0.007*"language"
Topic 10: 0.094*"thank" + 0.039*"thanks" + 0.036*"beautiful" + 0.031*"love" + 0.026*"much" + 0.024*"sharing" + 0.018*"great" + 0.017*"game" + 0.016
*"lol" + 0.013*"word"
```

Figure 10. top 10 topics

I got Figure 10 with additional stop words: removed, http, https, deleted, comment, post, rule. Words related to content deleted by violation of rules are no longer there.

Some topics listed can infer a relatively clear central idea. Topic 1 reflects a philosophical discussion about life and suffering. Conpared to topic 1, topic 3 delves into Buddhist understanding of self and the path to enlightenment. Topic 4 talks about Japanese Buddhism. Topic 5 seems to be talking about Buddhist practice of monks. Topic 6 deals with Buddhist precepts on not killing animals and eating meat. Topic 9 is related to Buddhist literature. Topic 10 is a cluster of gratitude comments, which is not topic specific.

Other topics are not so productive. Topic 2 and topic 8 are too vague to discern the idea. Topic 7 is confusing. It may be about physical representations like statues, but the presence of 'trauma' looks out of place.

I further experimented with the code by adding 'would', 'one', 'nice', 'cool', 'lol', 'also', 'thing', which turned out to be a too aggressive attempt.

```
Top topics with words:
Topic 1: 0.011*"open" + 0.010*"advice" + 0.008*"secular" + 0.007*"om" + 0.007*"parent" + 0.007*"freedom" + 0.007*"deeply" + 0.006*"year" + 0.006*"h
aha" + 0.006*"trauma"
Topic 2: 0.024*"christian" + 0.017*"buddhist" + 0.016*"head" + 0.016*"religion" + 0.014*"christianity" + 0.012*"drug" + 0.010*"thailand" + 0.009*"g
od" + 0.009*"jesus" + 0.009*"master"
Topic 3: 0.019*"tm" + 0.012*"buddhist" + 0.011*"india" + 0.011*"government" + 0.008*"monk" + 0.008*"hindu" + 0.008*"like" + 0.007*"guru" + 0.007*"y
es" + 0.007*"hinduism"
Topic 4: 0.057*"buddhism" + 0.049*"buddhist" + 0.034*"buddha" + 0.022*"teaching" + 0.016*"teacher" + 0.015*"tradition" + 0.013*"culture" + 0.012*"d
harma" + 0.011*"religion" + 0.011*"tibetan"
Topic 5: 0.021*"self" + 0.017*"buddha" + 0.015*"suffering" + 0.013*"mind" + 0.011*"view" + 0.010*"path" + 0.009*"body" + 0.009*"desire" + 0.008*"ca
use" + 0.008*"right"
Topic 6: 0.017*"people" + 0.013*"like" + 0.011*"think" + 0.008*"life" + 0.008*"way" + 0.008*"know" + 0.008*"time" + 0.007*"good" + 0.007*"even" + 0
.007*"get"
Topic 7: 0.050*"thank" + 0.025*"love" + 0.024*"thanks" + 0.022*"much" + 0.019*"beautiful" + 0.016*"great" + 0.013*"look" + 0.013*"sharing" + 0.011*
"like" + 0.010*"really"
Topic 8: 0.025*"book" + 0.022*"buddha" + 0.012*"read" + 0.011*"practice" + 0.011*"buddhism" + 0.009*"buddhist" + 0.009*"zen" + 0.008*"may" + 0.008*
"animal" + 0.007*"sutra"
Topic 9: 0.019*"people" + 0.016*"woman" + 0.012*"tattoo" + 0.010*"political" + 0.010*"authoritarian" + 0.009*"country" + 0.009*"trump" + 0.009*"whi
te" + 0.009*"gender" + 0.006*"american"
Topic 10: 0.013*"trump" + 0.012*"water" + 0.010*"minority" + 0.010*"war" + 0.009*"evil" + 0.009*"land" + 0.008*"pure" + 0.008*"man" + 0.008*"arise"
 + 0.007*"avalokiteshvara"
```

Figure 11. top 10 topics with too aggressive stop words

The new output is completely different. One the one hand, new topics emerged. Topic 2 is about cross-religion and cross-culture, and topic 9 and topic 10 are about politics and genders. On the other hand, topics on Buddhist precepts and Buddhist literature are missing.

Interestingly, the word ‘trump’ appeared several times in different topics. This weird phenomenon suggests the adjustment to stop words are too aggressive. I changed again the stops words to ‘removed’, ‘http’, ‘https’, ‘deleted’, ‘comment’, ‘post’, ‘rule’, ‘nice’, ‘cool, ‘lol’ and got the following results:

```
Top topics with words:
Topic 1: 0.020*"game" + 0.013*"play" + 0.012*"treatment" + 0.011*"name" + 0.009*"acceptance" + 0.009*"female" + 0.009*"sect" + 0.008*"forgive" + 0.
007*"youtube" + 0.007*"mass"
Topic 2: 0.027*"people" + 0.013*"feel" + 0.010*"like" + 0.010*"good" + 0.010*"person" + 0.010*"someone" + 0.009*"help" + 0.009*"compassion" + 0.008
*"want" + 0.008*"mental"
Topic 3: 0.014*"buddha" + 0.012*"one" + 0.010*"people" + 0.010*"thing" + 0.010*"like" + 0.010*"buddhism" + 0.009*"practice" + 0.009*"way" + 0.009*"
would" + 0.008*"think"
Topic 4: 0.043*"thank" + 0.019*"statue" + 0.018*"great" + 0.017*"beautiful" + 0.016*"thanks" + 0.016*"tm" + 0.016*"like" + 0.015*"buddha" + 0.014*"
look" + 0.013*"much"
Topic 5: 0.017*"would" + 0.013*"animal" + 0.009*"meat" + 0.009*"people" + 0.009*"like" + 0.009*"think" + 0.008*"woman" + 0.007*"control" + 0.007*"g
et" + 0.006*"man"
Topic 6: 0.052*"book" + 0.023*"monk" + 0.021*"read" + 0.017*"tattoo" + 0.011*"sutra" + 0.011*"pure" + 0.010*"lotus" + 0.009*"land" + 0.008*"one" +
0.007*"year"
Topic 7: 0.020*"love" + 0.017*"one" + 0.016*"may" + 0.011*"compassion" + 0.009*"happy" + 0.009*"anger" + 0.009*"peace" + 0.008*"life" + 0.008*"frie
nd" + 0.008*"wish"
Topic 8: 0.041*"buddha" + 0.029*"bodhisattva" + 0.018*"namo" + 0.015*"amitabha" + 0.014*"involved" + 0.010*"google" + 0.009*"deity" + 0.009*"shakya
muni" + 0.009*"tara" + 0.009*"search"
Topic 9: 0.032*"buddhist" + 0.029*"buddhism" + 0.009*"many" + 0.008*"country" + 0.007*"zen" + 0.007*"like" + 0.007*"culture" + 0.007*"also" + 0.006
*"religion" + 0.006*"group"
Topic 10: 0.034*"self" + 0.018*"suffering" + 0.018*"mind" + 0.015*"desire" + 0.015*"body" + 0.011*"view" + 0.009*"cause" + 0.009*"right" + 0.009*"p
ath" + 0.008*"karma"
```

Figure 12. top 10 topics with moderate stop words

Overall, I prefer the first version still. Some new topics appeared: mental health in topic 2, TM (transcendental meditation) in topic 4 and emotions and feelings in topic 7. But more confusing words that can be hard to associate with specific topics appeared: game, sect, youtube, google. Can I get rid of all of them? I do not think so. Actually, I learned that I should be careful about the choice of stop words. Only if the words are almost independent from the real content can they be deleted. For example, ‘removed’, ‘http’, ‘https’, ‘deleted’, ‘comment’, ‘post’, ‘rule’ can be seen as such, for they are seldom used when discussing topics related to Buddhism. Otherwise, ‘nice’, ‘cool, ‘lol’ can be used when discussing the topics to convey positive emotions. Even if they are not topic related, they can be used *together with* the topics. So when using topic models to detect possible topics, they are inseparable the topic related texts - they mix with topic related words when preserved, or break the boundaries between possible topics when taken away.

5. Conclusion and discussion

From the analysis above, the topics discussed under *r/Buddhism* on *Reddit* can be concluded as following categories:

a) Buddhist beliefs and concepts, such as self, suffering, mind, letting go, sentient beings, the four noble truths and the eightfold path.

b) Buddhist traditions, such as Zen and Japanese Buddhism, Tibetan Buddhism and Pure

Land.

c) Buddhist mantra, namely ‘om mani padme hum’.

d) Buddhist figures, such as guru and Thich Nhat Hanh.

e) Buddhist ethics on not killing animals.

f) Buddhist teaching, practice and literature in general.

Apart from the above outstanding topics, there are less prominent but possible topics as well:

g) Buddhist temples.

h) Cross-religion perspectives

i) Women in Buddhism and gender issues.

j) Political affairs.

One of the problems of this project is that there are not enough posts and comments. I only grasped 200 posts and the comments under them, so the topics discussed are very likely to be limited within 200 posts. However, there are much more posts available to obtain. If I want to get a more well-rounded look, I should increase the number of posts and comments.

It is a pity that the results lack personal experience, such as troubled relationships and personal experience in Buddhist practice like meditation (Actually, the word ‘meditation’ is not presented in any of the outputs, which is also surprising to me.). I think this also has something to do with limited number of samples. But more importantly, I think this is caused by the topic model I chose. The model generates a group of words that are associated with each other, but the topic is not given and should be concluded manually. When people talk about their lives, they are less likely to use terms or domain specific words. The words they use can be more general and more context-related, which makes them harder to detect. If I were to use another topic model that has a fixed set of topics, and a set of words that are linked to the topic, I might be able to detect topics related to personal life better.

Another problem is that there are a lot of images in posts. But the tools I used can only process the text in images. But images without text also have their meanings, like showing the beauty of a temple. Such ideas are missed in this project.

I also struggle in terms of whether the findings can yield informative results under the theoretical framework of religious studies. I feel like the current findings are mostly presenting, rather than explaining the subject. But I think there are possible interpretations or

hypothesis on the current findings:

a) If meditation is not frequently mentioned in the community, how people usually practice?

b) It seems that people can be involved in current political topics (Trump and animal ethics). How people conceive these issues in a Buddhist perspective?

c) Look deeper into the Buddhist teaching, practice and literature people are interested in to know about how people approach Buddhism.

The subject platform *Reddit* in this project can also be biased. It has more male, white and American users statistically, so the data collected can be biased in terms of overall representation. We should be aware that other online communities may have different results on the same research question. *Reddit* is an anonymous platform, and one user can apply for multiple accounts. This makes it hard to tell whether someone has posted a same topic multiple times. Although researchers using *Reddit* data as a source are becoming more and more popular, there are no systematic reviews of the contexts on *Reddit* that researchers are studying, nor the ethics practices they are engaging in relation to their work (Proferes et al., 2021), which indicates potential legal and ethical problems.

Finally, I have to make a statement about AI involvement. I used ChatGPT for writing and amending python codes, but the idea of how to process the data is originally created by myself. I would read and understand each step of the code before using the code. The results run by the codes are all evaluated by myself.

6. Reference

[1] Schmidt, T., Kaindl, F., & Wolff, C. (2020). *Distant reading of religious online communities: A case study for three religious forums on Reddit.* In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020)*. Riga, Latvia.

[2] Wikipedia contributors. (2019, March 6). *Reddit*. Wikipedia. Retrieved December 18, 2024, from https://en.wikipedia.org/wiki/Reddit

[3] Wikipedia Contributors. (2019b, November 10). *Thích Nhất Hạnh*. Wikipedia. Retrieved December 18, 2024, from https://en.wikipedia.org/wiki/Th%C3%ADch_Nh%E1%BA%A5t_H%E1%BA%A1nh

[4] Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society, 7*(2), 1–14. https://doi.org/10.1177/20563051211019004